



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

平成 30 年度 調査事業報告

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



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

事業名: 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

事業報告書

背景と目的	4
第一部 共通データベース化に向けた設計案	6
1. データベースとは何か？	6
2. 全国共通 DB 構築に向けたプロセス	7
2.1 Step1: 紙での管理(2019 年の児童相談所 DB の現状)	7
2.2 Step2: 標準的な DB	8
2.2.1 Step2-1: 情報共有用 DB の構築	10
2.2.2 Step2-2: 研究用 DB への自動データ整形	11
2.2.3 Step2-3: PoC に向けた実証実験と効果検証	12
2.3 Step3: 将来的なデータベース	13
2.3.1 Step3-1: 近接領域まで DB 共有範囲を展開する	14
2.3.2 Step3-2: AI によるリアルタイム意志決定支援	14
2.3.3 Step3-3: 記録の自動化や人材育成への発展へ	15
3. 必要なデータプラットフォームに向けた提案	16
4. DB 構築と運用に向けた前提条件	16
4.1 データ運用のポリシー	17
4.2 現場実務における情報共有 DB の条件	17
4.3 研究・評価のための研究用 DB の条件	18
4.4 政策・人員予算申請・業務管理への活用	18
5. 国際比較用のデータベースの具体的項目について	19
5.1 海外データベースについて	19
5.2 NCANDS のデータベース	19
5.3 NCANDS 子ども DB と児童相談所 DB の比較	19
6. まとめ	23
第二部 児童相談所データベースの項目例と活用可能性の検討	27
1. 目的	27
2. 方法	27
2.1 情報管理	27
2.2 データ抽出手続き	28
2.2.1 三重県データベースからのデータ抽出	28
2.2.2 神奈川県データベースからのデータ抽出	29
2.2.3 千葉県データベースからのデータ抽出	29
2.3 自治体 DB の項目整理	30
2.4 虐待再相談と対応期間の予測手法	30
3. 結果	32
3.1 児童相談所におけるデータベース項目の例	32
3.2 基礎集計	34
3.2.1 三重県の基礎集計	34
3.2.2 神奈川県の基礎集計	36
3.2.3 千葉県の基礎集計	37
3.3 ロジスティック回帰を用いた再相談予測と関連要因の検討	39
3.3.1 三重県における再相談予測モデルの結果	39

3.3.2 神奈川県における再相談予測ロジスティック回帰の結果	39
3.3.3 千葉県における再相談予測ロジスティック回帰の結果	40
3.4 対応期間予測と長期化要因の検討	41
3.4.1 三重県: 事例の対応長期化関連要因の検討	41
3.4.2 神奈川県: 事例の対応長期化関連要因の検討	42
3.4.3 千葉県: 事例の対応長期化関連要因の検討	43
3.5 機械学習による再相談事例と長期化事例の予測	43
3.5.1 三重県における再相談予測モデルの構成	44
3.5.2 神奈川県における再相談予測モデルの構成	44
3.5.3 千葉県における再相談予測モデルの構成	44
3.5.4 三重県: XGBoost による対応期間の予測モデル	45
3.5.5 神奈川県: XGBoost による対応期間の予測モデル	45
3.5.6 千葉県: XGBoost による対応期間の予測モデル	45
4. 考察	46
4.1 三重県の結果に対する考察	46
4.2 神奈川県の結果に対する考察	47
4.3 千葉県の結果に対する考察	47
4.4 リスクアセスメントの活用に関して	48
4.5 データベース設計への示唆	48
総合考察	49
第一部について	50
第二部について	51
まとめ	52
巻末資料	54
エグゼクティブサマリー	付録

背景と目的

虐待通告件数は増加の一途をたどり、平成 29 年度における全国の虐待通告件数は 133,778 件にまで及んだ(速報値)。相次ぐ虐待死の報道の中で、転居に伴う支援の停滞や関係機関間の連携不足が指摘されたこともあり、児童福祉機関間の情報共有や連携に対する社会的要請が高まっている。

関係機関が連携するにあたっては、その前提として情報共有が必要となるが、現状では、電話やメール、ファックスなどを使って属人的に行われている実態が指摘されている。増え続ける虐待通告に、関係機関が連携しながら対応するためには、共通のデータベースの設計が不可欠となるだろう。

増加する虐待通告に対して効果的な連携を実現するためには、伝達情報の抜けや漏れの無い体制を構築する必要がある。しかし、情報連携を前提としたデータベースの構想にあたって考慮すべき観点は多い。データベースに含まれる項目の内容については、従来のような「形式的な情報(e.g. 通告経路)」だけではなく、事例の内容的側面を端的かつ効果的に共有するための項目設計が求められる。システムの構成については、高い秘匿性が求められる個人情報をもどのようなネットワークで接続するかといった課題も伴う。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。データベースの目的と求められる機能について、改めて整理が必要であると言えるだろう。

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。一方、「活用」を想定したデータベース項目については、様々な設計の観点が必要となる。

活用を目的としたデータベースの主要な項目には、第一にリスクアセスメントに関する各項目情報が想定される。児童相談所における虐待対応の現場において、リスクアセスメントは支援の質を支える上での中核的な機能を有する。事例のリスクを適切に評価することによる死亡事例の防止や、児童の安全が継続的に脅かされる状態を防止するために、情報の蓄積が必要となる。過去のデータに基づき丁寧に精査されるべきであろう。

データベースとリスクアセスメントの関係については、「データの蓄積による知識の創出」という観点で接合される。年々積み重なる虐待事例とその対応結果、そしてその予後に関する情報が蓄積されれば、より良い対応へとつながりうる「知識」が生まれる。そして、過去の傾向から「どのような場合に、どのような対処を取ることが効果的か」といった予後予測と対応につなげることができる。事例の特徴を示すリスクアセスメントの情報をデータベースに蓄積することの意義は大きい。また、情報共有を前提とした際には、事例の重篤性や対応方針に関する連携機関間での共通言語としてアセスメント項目情報が機能しうるだろう。

既存の取り組みとして、国内では、奈良県・千葉県・神奈川県相模原市・岡山市・三重県などの自治体で、児童虐待に関係する機関の連携のためのデータベース作りに向けた動きがある。しかし、支援の質向上のための具体的な研究は未だ実施されていない。また、「活用可能な情報源」として、データベースにどの程度の情報量が蓄積されているかについても検討されてきていない。

本稿では、児童相談所における共通データベースの設計案について報告する。その前段階として、特にリスクアセスメントに関する項目に着眼した調査・研究結果をあわせて報告する。

第一部では、協力の得られた自治体のデータベース項目とその解析結果を援用し、「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への活用可能性がある項目などを調査し、その結果を整理する。

第二部では、協力が得られた3つの自治体(三重県・神奈川県・千葉県)のデータベースから、活用可能な知見を実例として抽出することを試みる。その際、各自治体のデータベース項目を概括的に整理した上で、(1)事例の予後はどのようなデータベース項目によって予測されるか、(2)データベースにはどの程度予後予測が可能になる情報量が含まれているか、の2点についてデータ解析を用いて検討する。

そして最後に、得られた結果を総括して、総合考察をまとめる。

第一部 共通データベース化に向けた設計案

1. データベースとは何か？

データベース(DB)とは、データを入れておくためのただの箱のことである。一方、データベース管理システム(DBMS)は、コンピュータのデータベースを構築するために必要なデータベース運用、管理のためのシステム、及びそのソフトウェアのことである。例えば、DBMSの機能には、データの入力フォームや検索機能などがある。一般的には、DBはほぼDBMSのことを指しているため、本章におけるDBはDBMSを含んだ意味とする。本章では、データベースの詳細には触れず、概要として知っておくべき内容をまとめる。

適切なDBを組むには、業務フローをおさえた上で、

-
- ①信頼性のあるDB
 - ②拡張性のあるDB
 - ③現場職員の方々のニーズに対応した設計
-

が必要になる。

①信頼性のあるDBとは、データがきちんとあるべきところに保存され、もしデータベースを置くサーバーになんらかの障害が発生した場合でも、迅速に復旧できることなどがあげられる。

②拡張性のあるDBとは、新しい変数の追加及び修正の容易さ、将来的に複数のDB同士をつなげられるような設計(例: 児童相談所だけでなく、市区町村や保健センター、医療機関、警察、学校・園などのDBとの連携)などが上げられる。

③現場職員の方々のニーズに対応した設計とは、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)するといった現場職員の方々のニーズを満たす機能である。

特に児童福祉現場では、法律改定や費用徴収計算の変更などに基づく新しい変数の追加(データ列の追加)があり、特に上記②、③については、現場の業務フローを把握していなければ、現場で使いやすいDB設計をすることは難しい。

児童相談所における既存のDBの多くは、各自治体がベンダー企業との保守運用契約の枠組みの中で、必要に応じて相談しながら、パッチワークのようにその都度継ぎ足しながら開発されてきた。そのこと自体は、各自治体が努力を重ねてきた結果だが、それゆえに大手ベンダーの児童相談システムでも、同じベンダーでも地域毎の営業所によってDB構造が全く異なっていることもあり、児童相談所DBは自治体ごとにガラパゴス化しているのが現状である。

2. 全国共通 DB 構築に向けたプロセス

プロセス概要(図 1-1)

Step1(2019年3月現在): 各自治体内でのみ運用される DB、紙での情報管理もある

Step2(2023年頃): 全国児童相談所をつなぐ標準 DB による情報共有

Step3(2027年頃): 標準 DB における AI(人工知能)及び RPA(Robotic Process Automation: 経過記録の自動化など)の活用

先に述べたように、児童相談所 DB は、各自治体毎に構造が全く異なっており、自治体間でデータフォーマットを組み合わせようとする、児童相談所設置自治体 69 カ所同士(実質的には導入しているベンダーの児相 DB の数)の組み合わせ問題となる。そのため、全国で統一した標準児相 DB を作るのであれば、まず国として標準 DB 構造を一つ設定し、それに各自治体 DB 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、DB 項目の調整を任せれば、局所的に整合性をとるような DB 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。

データベース (DB)整備に向けたプロセス



図 1-1: 今後の標準的な児童相談所 DB のプロセス

以下では、各段階のイメージを述べる。

2.1 Step1: 紙での管理(2019年の児童相談所 DB の現状)

2019年現在、多くの児童相談所で DB の実装・運用がされつつある。しかし業務の中では、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。あるいは経過記録は電子化していても、リスクアセスメントは全て紙、あるいは

は合計得点のみ電子化されており、各項目は紙にしか残っていないという場合も多い(図1-2)。

Step1：今現在



紙での記録
検索・保存・閲覧できない



FAX→メールでの情報共有
漏れが生じる
以前の関係機関情報が得られない



自治体毎にバラバラなデータ構造
統計カウントがバラバラ
情報共有ができない

図 1-2: 今現在の児童相談所データベース

メリット:

- ・DBを導入した自治体内であれば、記録管理できる。

デメリット:

- ・どのような情報を残すべきかについて、国として標準データベースの共通指針がないため、自治体ごとにバラバラなデータ構造になっている。
- ・自治体をまたいだ転出入の際の情報共有をDBとして想定していないため、情報共有の負担が大きく、また、共有される情報の質が担保されない(伝え手、受け手の技量に大きく依存する)
- ・全国調査対応などの際、現場で手集計する負担があり、またデータの正確性が担保されない

現場の虐待対応業務は、児童相談所だけで閉じるものではなく、多機関との情報共有が必要になる。実際、虐待死亡事例で自治体をまたいだ転出入が問題となることが少なくない。電話やFaxなど口頭やメモで情報を伝える際、伝える情報にバイアスが生じることはごく自然なことである。また、担当者の不在などで情報共有のために何度も電話することは、業務効率が非常に悪い。そのため、部門をまたいだ連携を行うためには、現場での情報共有に対応した標準DBの環境整備が必要不可欠である。

標準DBの整備ができれば、現場での情報共有だけでなく、厚労省調査や福祉行政報告例など統計調査への対応も、より現場の負荷が少なく、より正確に行うことができる。

2.2 Step2: 標準的なDB

Step1 の問題を解決するために、全国で統一した項目で情報を蓄積・共有できる仕組みの構築が重要である。

Step2 の標準的な DB の運用に向けたロードマップ

Step2-1: 情報共有用 DB を作る & API(Application Programming Interface の略で、データとプログラムの仲介役のような役割)によるアクセス権の制御(予想期間: 1 年)

Step2-2: 情報共有 DB から API によるデータ自動整形し、国際比較に向けた研究用 DB を作る(予想期間: 1 年)

Step2-3: Proof of Concept(PoC)に向けた実証実験と効果検証(予想期間: 1 年)

※Step2-1 から 2-3 まで同時に必要なこと: 業務フローの整理とデータ入力のマネジメント、及び人材育成・研修(予想期間: 3-5 年程度)

都道府県においても、構想から実装、そして実証実験まで 2 年～3 年程度の期間がかかることが予想される。虐待対応用の DB は基本的に 2 つある。現場における情報共有用の DB と、研究用の DB である。アメリカでは、1980 年代以後に虐待対応の政策決定に使うため研究用 DB から作り始めたが、コンピューター技術が進んできた現在、情報共有用の DB から作ることのほうが効率的と考える¹(図 1-3)。なぜならば、情報共有 DB ができると、そこから API などを通して研究用 DB はある程度自動生成できるからである。以下では、標準 DB の実装から実証実験までのロードマップについて述べる。

情報共有DBを作る

研究用DBは、情報共有DBからAPIを通して自動抽出する

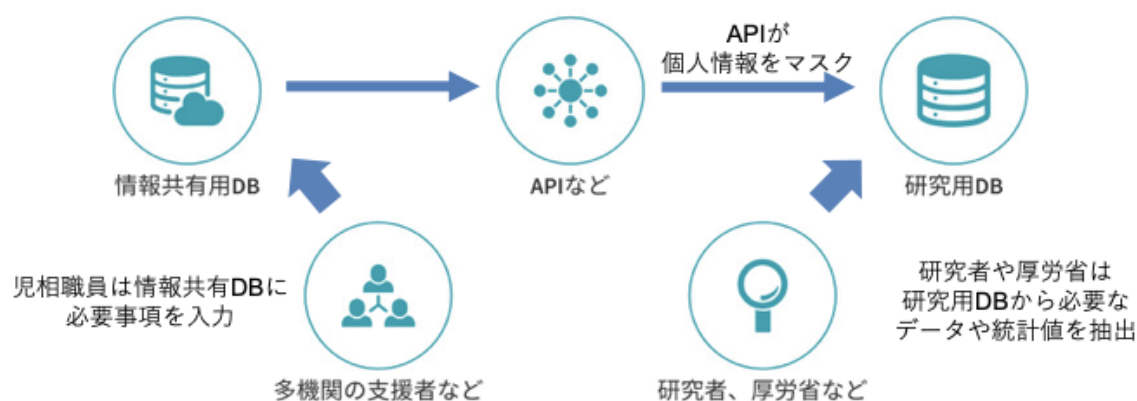


図 1-3: 情報共有 DB から研究用 DB への自動抽出イメージ

¹ 平成 29 年度 子ども・子育て支援推進調査研究事業「児童虐待の地域及び国際比較のためのデータベース構築に関する調査研究」報告書

2.2.1 Step2-1: 情報共有用 DB の構築

DB 設計と構築を考える上で、拡張性が高いことは非常に重要である。なぜならば、拡張性を考慮しない DB を作ってしまうと、現場の兎相 DB のように、必要に応じてパッチワーク的な継ぎ足し改修をすることになり、結果的にコストがかさむことになるからである。DB 構築のコストについて議論をするのであれば、導入時だけではなく、将来的な費用感も含めたシミュレーションが重要である。

またコスト観点だけでなく、パッチワーク的な継ぎ足し改修には、DB を複雑で使いにくい構造にしてしまう弊害があることも確認しておきたい。

拡張性を考慮すると、情報共有 DB は、本来は児童相談所だけでなく、自治体内の警察・医療機関・学校や園・保健センター・生活保護福祉事務所などの情報共有も視野に入れた構造にしておくことが重要である。具体的には、自治体内外の各組織のデータベースを、高度なセキュリティ、インターネットとつながらない閉域ネットワークで結ぶことにより、児童相談所で受理した事例について、関係機関に電話で情報を問い合わせるのではなく、1 クリックで必要な情報へのアクセス権申請と情報収集を行えるようなシステムである。

つまり、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(図 1-4)²。

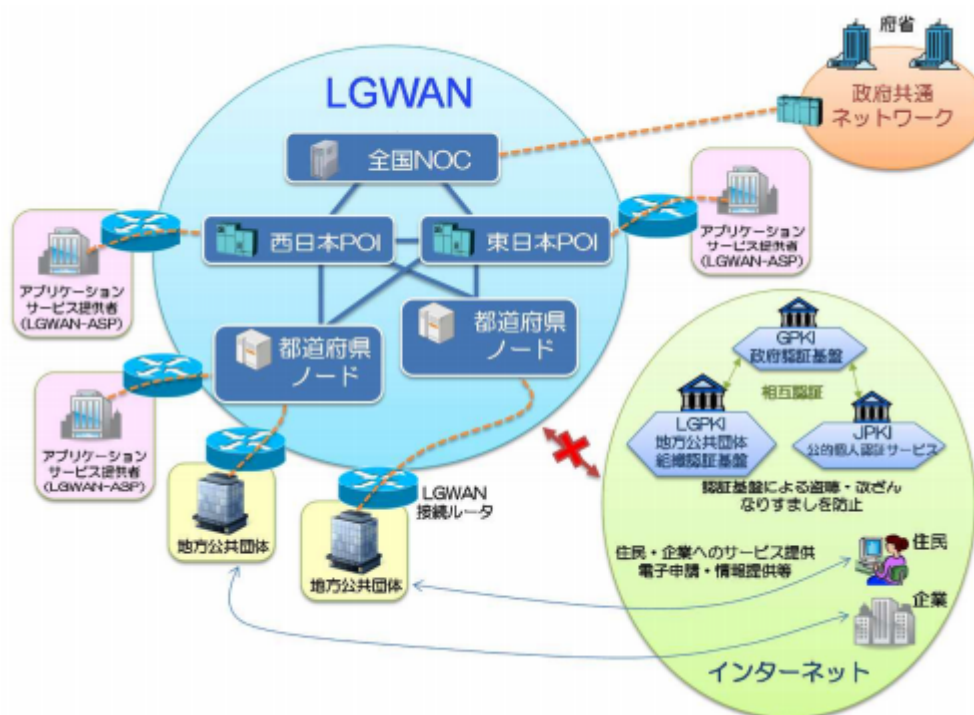


図 1-4: LGWAN 概要

拡張性についても LGWAN-ASP(Application Service Provider)の付加利用(有料)により、大容量 DB、高スペックな CPU/GPU など追加できるなど、拡張性についても備えている。そのため、既

² 総合行政ネットワーク(LGWAN)の概要 平成 30 年 4 月 地方公共団体情報システム機構 総合行政ネットワーク全国センター (https://www.j-lis.go.jp/data/open/cnt/3/180/1/L-2_gaiyou_Internet_201804.pdf)

に公共サービスとして LGWAN ではマイナンバー制度の利用や、災害時の情報共有などにも活用されている。児童虐待対応でも、全国の転出入における重篤事例の見過ごしを解決するためには、常時最新情報のフィードバックを共有できる LGWAN の利用を検討する段階にきている。

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 DB 共有ではなく、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用 DB を構築することが望ましい。

その上で、並行して時間を掛けながら自治体内の市区町村や警察、医療機関、学校・園などとデータを情報共有用 DB に保管することが求められる。情報共有といっても、全てのデータを筒抜けにするわけではない。各児童情報へのアクセス権の設定などは、中間サーバーの API と呼ばれるプログラムで管理・制御できるように設定し、どのケースにどの機関がアクセスしてよいかの設定をルールに基づき運用できるように構想することが重要である。

2.2.2 Step2-2: 研究用 DB への自動データ整形

API はアクセス権の管理だけでなく、自動データ抽出や整形なども随時行うこともできる。これができれば、情報共有用 DB に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ったり、研究者は自動整形された研究用 DB にアクセスできるため、現場へのアンケート調査を最小にすることもできる。情報共有用 DB から研究用 DB に情報を抽出・整形できる仕組みが構築できれば、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。

【情報共有 DB を設計する際に同時に構想すべき UI/UX】

情報共有 DB において、必要な情報を現場職員に入れてもらうためには、マネジメントと UI/UX の工夫が不可欠である。

iPhone や Android スマホも、バージョンのアップデートにより、使い勝手の向上とデザインの修正が行われているように、分かりやすいデザイン(User Interface: UI)や使い勝手(User eXperience: UX)を追求することは現代のテクノロジーにおいて非常に重要である。UI/UX にお金を掛けないと、新しい DB を作っても職員にデータを「入力してもらえない」、「使ってもらえない」。欧米でも、UI/UX の改善は DB 構築と同列の優先順位とされており、情報共有用 DB が機能するためには、必要な情報を入力してもらう UI/UX への投資は必要不可欠である。

そして、UI/UX を考えるとき、まずは入力デバイスを何にするかが非常に重要である。現状、児相の職員は、家庭訪問などの所外での調査や面接を行った後、一度、所に戻ってから記録を書いている。しかし、閉域ネットワーク利用、暗号化、仮想化、画面転送などの技術を利用すれば、外出先からセキュリティ高く DB に情報入力することができる。実際、民間企業は既にこういった仕組みを導入している。このような仕組みを利用すれば、タブレット PC 内にデータを溜めないため、仮にタブレット PC を紛失しても情報流出も起こらない。現在の紙での情報管理よりも、リスクが低いとも言える。

使いやすい画面デザインや入力の簡便さ(UI/UX)については、業務フローを分かった上で、UI/UX デザイナーとのデザイン構築が必要である。また、どのような入力項目が必要なのか、先行研究の知見と日本における過去データ、そして現場職員との意見交換も必要である。実際、三重県のリスクアセスメント事業ではどのようなリスクアセスメントの項目が必要なのかに加えて、どのようなデザインであればデータを入力してもらえるか、現場の児童相談所職員と UI/UX の意見交換に、2年半かけてブラッシュアップを繰り返してきた。

さらに一度作って終わりではなく、今後も現場職員によるデータ入力の簡略化や過去データから経過記録の入力補助機能など、使い勝手を向上させるために必要な機能を順次アップグレードしていく予定である。

現場においては、情報流出のリスクを最小限にすることが常に最優先されている。ただし、これだけ業務が逼迫している状況において、阻害されている業務効率化やAIを利用したリアルタイムシミュレーションや人材育成のベネフィットと、導入と運用のコストなども合わせて考慮し、総合的なバランスを考える必要がある。

「紙での管理が安全」と言いながらも、実際の運用では、訪問先の住所を紙にメモしたり、庁有車のカーナビやGoogleマップで直接住所を入力することは行われており、そのデータを消さない、メモや児童票コピーを入れた鞆を置き忘れる、盗まれるというインシデントは既に現場で起きている。そのような状況に対して、「紙での管理が安全」という認識自体、事実に基づいて更新する必要がある。

以上のように Step2-2 で必要な作業をまとめると、図 1-5 となる。

Step2：標準的なデータベース



図 1-5: Step2 標準的なデータベースを作る際の手続き

2.2.3 Step2-3: PoC に向けた実証実験と効果検証

予算が限られる自治体において、単一年度で完璧なDBを仕組みとして構築することは難しい。そのため、現場のニーズに基づき、優先順位を決めながら、小規模に実証実験と効果検証をすることが求められる。また大都市部と地方部では、抱えている課題や前提となる環境が異なるため、各地域における特徴を吟味した効果検証が求められる。2019年5月からは三重県児童相談所での実証実験がスタートする予定である。

【Step2-1 から 2-3 まで同時に必要なこと：業務フローの整理とデータ入力のマネジメント、及び人材育成・研修】

人材育成&研修で必要なこと

- ・データ入力の数カントルール
- ・DB 入力に関する定着に向けた研修
- ・データに基づく人材育成

Step1 で述べたように、各自治体のデータ入力のルールは標準的な運用方法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り異なっている。³

例えば・・・

- ・3人きょうだいが受理される際、家族単位で1ケースとして受理する自治体と、3ケースとして受理する自治体
 - ・年度が替わった際に、新規ケースとして取り直すDB構造
 - ・そもそも業務が忙しく、データを入れる時間がない
 - ・担当者により、必須データ入力項目にもデータが入っていない
- など、標準的なカントの問題、データベースの問題、組織要因、担当者要因などがあげられる。

特に、今後リスクアセスメントを導入する際には、研修を継続して行わなければ、誰が入力したかによって、リスクの傾向が異なってしまう。そのようなブレを少なくするには定期的な研修や人材育成を前提にしたリスクマネジメントが必要である。言い換えると、リスクアセスメントをシートとして導入しても、研修体制や模擬事例に対する担当者毎のリスクに対する視点の違いなどを議論できる人材育成体制がないと、すぐに形骸化しデータとしての価値も失われてしまう。そのためには、マネジメント体制として、常にリスクアセスメントに関する研修、事例検討会、管理職研修などをした上で、リスクアセスメントを運用することが重要となる。

2.3 Step3: 将来的なデータベース

Step3 の目標は以下のような目的である。

Step3-1: Step2 の情報共有 DB を近接領域(例えば、医療機関・警察・母子保健・生活保護・学校・園など)と結合・実装(調整期間: 5-10年、実装予想期間: 2-3年)

Step3-2: 蓄積されたデータをリアルタイムで AI に学習させ、先達の知見とデータに基づく新たな知見をケースに参照・活用する(予想期間: 3-5年)

Step3-3: 経過記録を溜め、記載内容をパターン化することで、RPA(Robotic Process Automation)により、経過記録の自動生成、音声認識による簡易な記録生成プログラムなど(予想期間: データサンプル次第)

以下に将来的なデータベースのイメージをまとめる(図 1-6)。

³平成 27~28 年度 全国児童相談所長会委託定例調査 虐待通告の実態調査(通告と児童相談所の対応についての実態調査) 報告書 (山本・高岡・久保・坂本, 2018)

Step3：将来的なデータベース



図 1-6: 将来的なデータベース活用イメージ

2.3.1 Step3-1: 近接領域まで DB 共有範囲を展開する

世界で最も自治体のデジタル化と情報共有が進んでいるエストニアは、セキュリティの高い暗号化されたネットワークにより e-Government 化を進め、あらゆる公的機関の DB を結んでいる。虐待事例の場合は、児童相談所に通告があった後、保健センターのような訪問記録、保育園・幼稚園での行動記録、学校での成績や出席情報、飲んでいる処方薬情報などを、ほぼ 1 クリックで情報収集できる。その結果、情報共有にかかる時間はほぼ秒単位で済み、現場職員は、子どもと家族への対応にほとんどの時間とエフォートを割くことができているという⁴。

Step3-1 では、虐待対応だけでなく、アクセス権の制御とセキュリティ設定した上で、関連機関との大規模ネットワーク化を進めることが望まれる。Step2 においては全国児童相談所をつなぐことを提案したが、Step3 では、さらに児童相談所から各自治体に多くのデータをつないでいく段階である。具体的には、Step2 でも述べたように、医療機関や警察、保健センター、学校・園、福祉事務所など、必要な情報を 1 クリックで収集できるシステムである。仕組みとしては LGWAN にデータを保存し、かつ LGWAN 上のネットワークでデータベースを連携するイメージとなる。

2.3.2 Step3-2: AI によるリアルタイム意志決定支援

人工知能(AI)は、ビッグデータ化、機械学習の発展、そして計算機の高速化が進んできたことで、精度を上げてきた。しかしながら、Step1 で述べたように、現状の標準化されていない児童相談所 DB では、全国のデータを結合することもできないし、それをそのまま解析することもできない。だからこそ標準 DB が必要となる。

標準 DB ができ、データを結合できるようになると、AI の大規模活用の可能性が見えてくる。もちろん、データは全てを解決するわけではないし、AI も万能ではない。ただし、

⁴ eGA “デジタルトランスフォーメーションと e-Government” 視察 28 Jan. 2019- 3 Feb. 2019

現場でどう対応して良いか分からない時は、現場では先輩や同僚に相談するように、AI自身に学習させて、どのような特徴があるケースはリスクなのか、あるいはどのような特徴があれば早期に終結しても再発確率が低いのかを現場判断の中でリアルタイムに参照することができる。そのため、業務におけるデータを集めることは、そのままAIの現場活用に繋がっていき、記録の保全だけでなく、どのような対応をすると上手く支援がいくのか、あるいはどのような対応をすると支援が失敗するのかといった知見を、すぐ次の事例に活用できる好循環を生み出すことにつなげることが可能となる(図1-7)。即ち、児童虐待対応の多機関連携を含めて解析は、ここまで来ることでやっと可能になり、かつデータの利活用が業務と連動させることができる。

①業務記録→

②知見のデータ化→③AI学習→

④知見を次ぎケースに活用→⑤記録



図1-7: 記録、データ化、AI活用、知見の活用のプロセス

2.3.3 Step3-3: 記録の自動化や人材育成への発展へ

今後の児童相談所の業務効率化を考えた場合、記録の効率化があげられる。効率化については、2つの可能性がある。一つは、音声入力による記録手法である。ただし、2019年2月現在音声認識と自然言語処理などにより、インターネットに接続することができればある程度の精度を持った記録は可能になっているものの、個人情報を含む機微情報についてはインターネットに触れることは推奨されない。今後、オフラインでの利用で精度が保たれるレベル、またはLGWAN-ASPなどでGoogleやAppleなどの音声認識レベルが運用上担保されない限り、現段階での利用は難しいと考えられる。一方で、データが蓄積されてきた時に、どのようなケースであればケース記録を簡略化、または自動化(RPA: Robotic Process Automation)できるかなど、データに基づき判断することができるだろう。

また、データは人材育成や研修でも活用できるメリットがある。一つは、AIにベテランの知見を学習させることができれば、また人が異動や退職などで変わっても、先達のベテランの知見を引き継げる。さらにAIは機械だからこそ、24時間365日、いつでも活用することができる。二つ目は、より職員の特性にあった研修に活用もできるだろう。児童相談所

の研修体制は、新人についてはリテラシーや知識の均等化のために同一の研修を行うことは大切と考えられる。しかし、中堅職員以上になった時、その時点のスキルやデータを職員自らが把握できれば、必要なスキルやマインドセットを職員自らがカスタマイズして受講したり、個々人の支援者特性に基づくキャリアアップに活用できると考えられる。なぜならば、子どもの個別学習にとっては得意不得意科目に応じて、パーソナライズされた教育研修体制が推奨されているのに、大人の児童相談所職員にはそれを適応せず一律研修というのは無理がある。行政の無謬性においては、誰が担当しても間違わず、均一なサービスをという名目となるが、対人支援においては支援者の特性と子どもや家族の特性のマッチング問題はごく当然存在するものである。そのような支援者の個性を否定するのではなく、ケースワークに上手く活用していくためにも、データは効果的な支援の質の向上に活用する上でも必須インフラとなる。

3. 必要なデータプラットフォームに向けた提案

上記のステップを鑑み、標準 DB 構築だけでなく、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(図 1-8)。なぜならば、標準 DB は全国の児童相談所、及び関係機関をつなぎ、情報共有ができることが重要である。ここでいう情報共有とは、CA 情報のように Excel シートだけを自治体間で共有するのではない。現場に必要な情報は、絶えず更新される最新情報であって、一度きりの情報共有ではないからだ。言い換えると、CA 情報では基本的に一度出したら終わりの単回性の性質であるのに対し、現場で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。それを実現するには、LGWAN のような各自治体をつなぐネットワークの利用を前提に考えることが重要である。

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



図 1-8: 必要なデータプラットフォームに向けた開発事項

4. DB 構築と運用に向けた前提条件

前節では、ステップとして必要な内容をまとめたが、本節では具体的に Step2: 今後の標準 DB を設置するときに事前に検討すべきことについてまとめる。

標準 DB 構築の内容について、まずはデータ運用のポリシーを 4.1 にまとめる。それを前提に 4.2 現場実務における情報共有 DB の条件、4.3 研究・評価のための研究用 DB の条件について述べる。そして 4.4 政策・人員予算申請・業務管理への活用についてまとめる。

4.1 データ運用のポリシー

・データ及び解析結果は、あくまで参照事例であって、最終的な決定は人が決定する。

現場のケースワークにおいて、「もし現場判断が間違っていたら事例を振り返り(検証し)、課題のある対応を修正する。良い対応は、さらにそれを強化する」というアクティブラーニングのサイクルは、現場支援者であっても、AI における学習モデルの更新の視点でも同じである。そのため、AI によってケースワークが振り回されるのではなく、あくまでデータの分析結果を参照しながら、データと現場の知見を融合するような AI 運用ポリシーが必要である。

4.2 現場実務における情報共有 DB の条件

現場実務に必要な条件の詳細については以下の機能や DB 設計が必須と考えられる。

-
- ①使いやすいフロントデザイン
 - ②福祉行政報告例の自動エクスポート機能
 - ③リスクアセスメントは合計得点ではなく、各項目毎の保存
 - ④リスクアセスメントを複数回に分けて取った場合は、リスクの遷移を見られることが重要なため、各得点を取った回数分全て残す
-

①業務で使いやすいことが前提: 児童相談所職員が使いやすく、かつ入力しやすいものとして、業務フローに合わせて良く練られたフロントデザイン(アプリの使いやすいデザイン)が重要である。また、関係機関における情報共有を前提としたものであり、かつセキュリティが高いネットワークとクラウド DB の併用が必要として LGWAN の利用が重要である。特に、自治体間での情報共有には、各自治体共通のフォーマットが必要であることから、そのフォーマットは厚労省が国の標準仕様として準備・通知することが求められる。

②児相事務業務を簡略化するため、福祉行政報告例に合わせた自動エクスポート機能などが不可欠である。

③リスクアセスメントは信頼性・妥当性が検証されたものでないと、解析的に意味のある結果を出すことができない。そのため、信頼性・妥当性があるリスクアセスメントへの改変が必要である。また、特に、リスクアセスメントは合計得点よりも、各項目の得点のほうが重要なデータとなる。データベース設計において、データベースの合計得点だけを残すような設計は解析的にすべきではない。

④リスクアセスメントは最終結果だけでなく、緊急受理時、現認時、援助方針会議事、終結時など、最低 3 つ、4 つのポイントにおける時系列のリスク変化を見ることが大切である。

1章でも述べたように自治体DBは、最終時のリスクアセスメント合計得点しか入っていない場合が多く、それだと予測精度がかなり落ちてしまう。そのため、データの分析やAIによる知見を検討するには、各事例がどのようなリスク特徴があるかを把握するために、リスクアセスメントの合計得点だけでなく、各リスクアセスメント項目のデータを、取得時のタイムスタンプと共にDBに保存することが重要である⁵。

4.3 研究・評価のための研究用DBの条件

研究・評価のために最低限備えておくべき機能は以下と考えられる。

-
- ①年度をまたぐ事例のアンケート調査や複雑なデータ抽出をするために、一括エクスポート機能を持つこと
 - ②国際比較するためのデータ形式へAPIを利用した自動抽出及び整形機能
-

①各児童相談所から知りたいデータを取り出したり、大学や国立研究所、データサイエンス民間企業などに依頼する場合、一括エクスポート機能を実装しておくことが重要である。

②現場の情報共有用の標準DBを活用し、そこから研究用DBを作ることを基本機能として備える必要がある。研究用DBの管理者は国としての集約で取りまとめが必要なため、管理は厚労省、または大学・国立研究所などが候補になると考えられる。

4.4 政策・人員予算申請・業務管理への活用

以下では、情報共有DBと研究用DBを用いて、どのように政策・人員予算請求・業務管理に活用できるのかについて述べる。

-
- ①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
 - ②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
-

国際比較ができるDBが準備できると、同じようなリスク事例に対して、海外だといくらぐらいの予算をつけているのか、またどのような人員体制をつけているのかを、データに即して提案することも可能となる。また、虐待対応件数の予測データから人員予算請求の基礎資料を作成することもできる。

また、児童相談所の効率化は、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、子ども一人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、地域における犯罪

⁵ もし既存の自治体DBについて、一番最後のリスクアセスメントデータしか残らず上書きされるのであれば、物理的に古いデータを削除するのではなく、DB上に削除フラグ(0/1)を立てて論理的に削除したかのように見せるような仕組みもある。

率や貧困率などを踏まえたケースの困難度(リスク)や児童相談所が管轄するエリアの広さ(一時保護所までの移送時間)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。

5. 国際比較用のデータベースの具体的項目について

5.1 海外データベースについて

昨年度の海外 DB 報告書として、全米児童虐待及びネグレクトデータシステム(NATIONAL CHILD ABUSE AND NEGLECT DATA SYSTEM: NCANDS)について報告した。その他、世界的な国際比較用データベースは国際虐待防止学会が取りまとめを行う ICAST などもあるが、北米における NCANDS のデータセットは構造的で、カナダもそれに準じてデータベースを構築しているため、日本の国際比較を目指すためには、第一にこの NCANDS のデータ構造に合わせた形式が望ましいと考えられる。

5.2 NCANDS のデータベース

NCANDS のデータベースには、二種類ある。

- ①福祉行政報告例のように組織や自治体単位における虐待対応件数などを報告するもの
- ②個人情報がマスクされた上で概要を把握できるものがある。

①については、アメリカにおける虐待通告に対するスクリーニング制度の違いがあるため、①のみで詳細な国際比較をすることは難しい。しかし、②においては、ある程度網羅することができると考えられる。また、②を満たすことができれば、そこから①に対応したデータがある程度自動抽出することが可能と考えられる。

そのため、本節では②の子どもの概要把握データベースについて議論を進める。②の子どもの概要把握データベース例については巻末資料1に示した。

5.3 NCANDS 子ども DB と児童相談所 DB の比較

NCANDS の主要項目をまとめ、三重県、神奈川県、千葉県 of 既存の DB 構造を比較する。下図において、左列が NCANDS の主要項目、右列が三重・神奈川県・千葉県の既存 DB の項目である。

NA(Not Available)の部分のみ、現状日本における児童相談所 DB でデータ化されていない項目となる。表内の「NA: ただし DB にはアリ」については、現状の DB にも入っているので、そこだけを項目として抽出するプログラムを書けばよい。一方、「NA: ただし DB 児童票に記載アリ」及び「NA: ただし DB 経過記録に記載アリ」については、担当者によっては児童票または

経過記録に記載している内容と考えられる。この部分は現場の進行管理としても必要な項目であることから、新たに標準的データベースとして厚労省からの通知等により、標準 DB に追加することができれば対応可能と考えられる。

表 1-2: NCANDS と三重県/神奈川県/千葉県の DB 比較

NCANDS 各子ども事例	三重県/神奈川県/千葉県の DB 利用項目
1. 児童虐待事例の発生年度	年度
2. 児童虐待が発生した場所	居住市町
3. 児童虐待の報告書ごとに割り当てられる ID 番号	ケース番号
4. 児童ごとに割り当てられる ID 番号	児童番号
5. 児童相談所が児童虐待の疑いの通告を受けた日時	保護判断時通告日
6. 調査開始年月日	期間データ_児童受付日
7. 虐待の疑いを通報した人物・機関のカテゴリー	受付経路
8. 児童虐待に対する児童相談所としての判定	保護判断時安全に対する意思決定
9. 児童相談所によって判定が下された日時	一時保護_決定日
10. 同一事例に対する複数の関係機関からの通告有無	≒過去の係属歴
11. 被虐待児童の年齢	児童年齢
12. 被虐待児童の生年月日	NA: ただし DB にはアリ
13. 被虐待児童の性別	児童性別
14. 被虐待児童の国籍	NA: ただし DB にはアリ

15.	被虐待児童が居住していた地域	居住市町
16.	虐待が疑われる事件時の児童の居住環境	NA: ただし DB 経過記録に記載アリ
17.	被害歴	受付区分
18.	虐待のタイプ	主たる虐待種別
19.	虐待の判定	保護判断時一時保護の実施委託含む
20.	虐待死	NA: ただし DB 児童票に記載アリ
21.	児童自身のアルコール依存の問題	NA: ただし DB 児童票に記載アリ
22.	児童自身の薬物依存の問題	NA: ただし DB 児童票に記載アリ
23.	児童自身の知的障害の問題	保護判断時精神遅滞
24.	児童自身の精神障害の問題	保護判断時診断あり
25.	児童自身の視覚障害または聴覚障害の問題	NA: ただし DB 児童票に記載アリ
26.	児童自身の学習障害の問題	NA: ただし DB 児童票に記載アリ
27.	児童自身の身体疾患の問題	保護判断時身体障がい
28.	児童自身の素行の問題	NA: ただし DB 児童票に記載アリ
29.	児童自身のその他病状に関する問題	NA: ただし DB 児童票に記載アリ
30.	保護者のアルコール依存の問題	NA: ただし DB 児童票に記載アリ
31.	保護者の薬物依存の問題	NA: ただし DB 児童票に記載アリ

32.	保護者の知的障害の問題	NA: ただし DB 児童票に記載アリ
33.	保護者の精神障害の問題	NA: ただし DB 児童票に記載アリ
34.	保護者の視覚障害または聴覚障害の問題	NA: ただし DB 児童票に記載アリ
35.	保護者の学習障害の問題	NA: ただし DB 児童票に記載アリ
36.	保護者の身体疾患の問題	NA: ただし DB 児童票に記載アリ
37.	保護者のその他病状に関する問題	NA: ただし DB 児童票に記載アリ
38.	家庭内暴力	NA: ただし DB 児童票に記載アリ
39.	不適切な住居環境	NA: ただし DB 児童票に記載アリ
40.	児童の家族の経済的問題	NA: ただし DB 児童票に記載アリ
41.	公的支援	NA: ただし DB 児童票に記載アリ
42.	虐待判定日から 90 日以内に行われた介入	NA: ただし DB 児童票に記載アリ
43.	介入を開始した日時	≒措置データ_措置開始日
44.	社会的養護	NA: ただし DB には児童票に記載アリ
45.	社会的養護へ委託された日時	≒措置データ_措置開始日
46.	担当者 ID	主担当者番号
47.	スーパーバイザーID	NA: ただし DB にはアリ
48.	加害者 ID	NA: ただし DB にはアリ

49.	加害者関係性	主たる虐待者
50.	加害者の親としての立場	NA: ただし DB にはアリ
51.	加害者のケア提供者としての立場の有無	NA: ただし DB にはアリ
52.	通告時の加害者年齢	保護者年齢
53.	加害者の性別	保護者性別
54.	加害者の国籍	NA: ただし DB にはアリ
55.	加害者の虐待加害の前歴	NA: ただし DB にはアリ
56.	「加害者」との「18.虐待のタイプ」との関係	NA: ただし DB にはアリ
57.	事件日	NA: ただし DB 児童票に記載アリ
58.	通告時刻	NA: ただし DB にはアリ
59.	調査開始時刻	NA: ただし DB にはアリ
60.	死亡日	NA: ただし DB にはアリ
61.	社会的養護の終結日	期間データ_児童終結日

また、日本の標準 DB として日本独自に持つべきデータは、警察からの通告(相談種別、相談種別詳細、通告者、通告者詳細、身柄付き通告)、泣き声通告、DV 通告、過去の DV 係属歴、特定妊婦、いじめ相談、児童買春等被害相談、1 歳 6 ヶ月健診受診状況、3 歳健診受診状況、転出入歴などは最低限持つべきと考えられる。

6. まとめ

児童相談所の標準的 DB を作る際には、現場の情報共有 DB と研究用 DB があり、まずは現場で使いやすい業務用 DB を作り、そこから自動的に研究用 DB として抽出できる形式が

望まれる。これらは昨年度報告書において、アメリカの CCWIS の形式でデータを集めることの形式に近い⁶

昨年度の報告書で示した内容を以下に再掲する。

1355.52CCWIS ルール

“1355.52 CCWIS project requirements(<https://www.law.cornell.edu/cfr/text/45/1355.52?qt-ecfrmaster=3#qt-ecfrmaster>)”

という全ての州が遵守する運用ルールがある。データは日々収集され、レビューや監査も同時並行で行っていく。運用基準の最も重要な点は、データをどのように国に集めて、その質を保つか、そしてデータの質の管理を守ることに関わる。そのため、情報交換の仕方についても運用ルールを規定している。できる限り正確で一定の質を守るために、全ての子どもの福祉に関する情報を統括し、維持していく必要がある。そのためにも組織間での貢献が重要になってくる。メディケードや、虐待・ネグレクトの情報、母子に関する情報、法律、裁判所の申請状況など、それらの情報をどのように共有していくかの基準を設けたことで、電子的にデータ共有が非常にスムーズに運用されるようになった。子どもの福祉に関する関係組織が CCWIS をユーザーとして活用しており、それに付随するシステムを子ども局が構築している。なお、運用基準に関しては子ども局が独自に作成している。

基本的に情報共有用の DB については基本システムはアメリカでは子ども局が構築し、かつ運用基準も子ども局が設定している。日本においても、標準 DB となる情報共有 DB について、国としての舵取りが必要不可欠であると言える。このような標準 DB の形態を自治体間の調整でさせることは無理がある。

また、DB を作る際には、都道府県を越えた児童相談所間の情報共有、及び児童相談所と警察や医療機関、学校・園など関係機関との情報共有も DB 拡張性に組み込むことが必要である。拡張性を考えなければ、コスト的に自治体にとってベネフィットが見合わない。

実質的には、アメリカではコストの問題がある。昨年度の報告書では、費用の問題については以下のように記載されている。⁷

90年代から継続して開発を進めてきているし、当時の時代背景や貨幣価値などからいくらかという算定は難しい。また各州との連携もあり、運用基準が幅広いので、どういうものを構築させるかによって変わってくる。ただし、設置・導入の際に **10 億円-100 億円** 程度の間となる。その後、保守運用費用・アップデート費用、通信費なども発生する。

⁶ 平成 29 年度 子ども・子育て支援推進調査研究事業「児童虐待の地域及び国際比較のためのデータベース構築に関する調査研究」報告書 13 ページ

引用: <http://www.boshiaiikukai.jp/img/research/houkoku-kodomokosodate-29.pdf>

⁷ 1 の報告書内 16 ページ

としてまとめられている。国としての方向性を確認し、同時に最低でもアメリカとの国際共同研究を実現する場合、少なくとも10億円から100億円単位を想定することが必要である。

一方、費用感の問題として、以下のようなデメリットも昨年度報告書に記載されている。

8

多くの課題が存在する。費用負担面が大きい。地方のレベルになるとそのための予算を組むことが問題となる。新しいシステムに移行できないと、古いデータのままとってしまう。新しいものに対応できなくなり、場合によっては古いデータはどこかにいってしまうことにすらなる。新しいシステムと古いシステムをどのように共有していくかが重要である。例えば、それぞれデータに関するファンディング・リソースが州によって違うが、予算がつかなければ、実際の古いデータが取れなくなってしまうこともあった。複数の有用なシステムを違った方法で使えるように郡や県のレベルで構築していかなければならない。地方の構造によっても異なる。良いアイデアがあっても予算がなくてはできない。予算ができた時にその都度システム改修を行い、導入していくということも大事ではある。

以上のように、現場のデータをまとめるという点において、一定のお金が必要不可欠であり、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统一で示すという国主導のスタンスがどうしても必要となる。厚労省は、平成31年度に虐待防止のための情報共有システム構築事業として、1自治体あたり4000万円(1/2補助)を方針として打ち出している⁹。具体的には、「新たに都道府県も実施主体に加えた上、市町村の関係部署や児童相談所(都道府県)等の関係機関間のより効率的な情報共有を進める」と記載されているが、転入・転出を含む自治体間における情報共有でなければ、死亡事例に対する施策としては片手落ちになってしまう。

さらに、各自治体で閉じた情報共有DBとは、データ構造がベンダーによる各自治体毎の独自ルールでデータベースを作ることになるため、自治体間を越えた情報共有は全国统一で動けなくなる。そのため、情報共有DBとして拡張性を考えた場合、自治体内に閉じた形でデータを運用するのではなく、都道府県をまたぐ事例を前提に、現場では常に最新情報に同期された仕組みを構築しなければ、本来の意味でのベネフィットは見込めない。そのための方向性として、最初から現場の情報共有DBを、自治体間をつなぐネットワークであるLGWANに乗せ、全国的にセキュリティが高く、かつ随時最新情報のフィードバックが来るような情報システムを方針として打ち出すことが重要である。

事実、社会的養育ビジョン¹⁰によれば、データベースについて

⁸ 1の報告書内16ページ

⁹ <https://www.mhlw.go.jp/content/000352155.pdf>

¹⁰ <https://www.mhlw.go.jp/file/05-Shingikai-11901000-Koyoukintoujidoukateikyoku-Soumuka/0000173888.pdf>

・児童相談所及び市区町村で使われている業務統計を見直し、現状に合っている形で定義を明確にした統計に変更するため、変更案とその定義を提示し(平成 29 年度)、一部の地域で新統計をモデル的に試行し(平成 30 年度から)、全国での統計を変更する(平成 35 年度)

【国】

・特に、虐待統計に関しては、海外との比較も行えるよう、児童相談所と市区町村への通告件数等、統計項目を検討して定義を明確にし(平成 29 年度)、一部の地域でモデル的に試行し(平成 30 年度から)、新たな統計を開始する(平成 35 年度) **【国】**

・一時保護及び代替養育や在宅措置を行った子どものフォローができ、プライバシーが守られるようなデータベースの構築を検討(平成 29 年度)し、そのデータを集めて検討できる研究が行える機関に集約できるシステムを構築(平成 33 年度を目指す) **【国】**

・児童相談所、市区町村の子ども家庭福祉、一元化された通告窓口などが通告情報を共有できるデータベースを構築して、都道府県に提示 **【国】** (平成 34 年度)

・上記データベースを用いた都道府県の対応を開始する **【都道府県】** (平成 35 年度)

と述べられている。国としての社会的養育ビジョンにおける標準データベースを目指すには、全国の自治体のデータ構造をまとめることが不可欠であり、明らかに情報共有 DB と研究用 DB を同時につくりつつ、かつ都道府県をまたぐ転出入も含めたフォローができる(特に児童相談所が関わった事例)データベースの構築が必要となる。これらは都道府県に

【国】が提示となっており、平成 30 年度から一部の地域でモデル的に試行し、平成 34 年度までに都道府県に **【国】**が提示となっているため、国としてのデータの標準化を進める必要性が高い。

そのようなデータの標準化を自治体に任せると、自治体同士の地位は公平であり、かつ自治体毎に情報セキュリティの基準やデータの構造が違うので、それを調整するのは時間とお金だけが掛かってしまう。さらにいうと、自治体の児童相談所職員は子どもの安全を守る専門家であって、データベースの専門家ではない。DB が専門ではない児童相談所職員にそのような専門外の重要な判断を任せるのではなく、専門性を活かせるように議論を進める必要がある。

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。

第二部 児童相談所データベースの項目例と活用可能性の検討

1. 目的

記録としての側面ではなく、データベースの「活用」可能性を実践的に検討するため、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行う。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。また、標準的なデータベースの構想を目的として、3つの自治体のデータベース項目を総合的に整理する。

このような目的を達成するためには、個人情報の扱い方や、データベースからのデータ抽出についても整理する必要がある。次節では、(1)情報管理、(2)データ抽出手続き、(3)データベースの項目整理方法、(4)本項で利用する統計解析手法についてまとめる。

2. 方法

2.1 情報管理

本調査では、個人情報記録されたデータベースからのデータ抽出を行い、抽出されたデータを受託機関(産業技術総合研究所)内で取り扱うという形式を採用した。個人情報は匿名化された状態で受け渡す形式を採用。具体的には、氏名、住所情報をあらかじめ除外し、氏名については解析用ID番号を付与することで匿名化を行っている。ただし、データ提供自治体については、自治体側がID番号と実事例番号との対応表を保持しうる形式とした。そのため、解析データ自体は匿名化されているが、対応表との結合が万が一実現された場合には個人情報と同等の秘匿性を有していることとなる。また、協力の得られた自治体の一つ(神奈川県)では、データベースに含まれる児童相談対応関連項目が含まれたデータのバックアップファイルを暗号化してハードディスクに保存し、パスワードロックをかけて預かる形式を採用した。バックアップデータには、対応経過記録テキストなど、個人情報を含むデータが含まれている。

各自治体から預かったデータは、全て鍵のかかる保管庫に保管された。ハードディスクに保存された元データは、ハードディスク自体にかけられた電子パスワードに加え、3重のセキュリティロックがかかる産業技術総合研究所内の解析室に置かれた鍵のかかる保管庫にて保管された。解析時には、同様の3重セキュリティロックがかかる解析室内に設置された外部通信が一切遮断されたスタンドアロンPCを利用し、そのPC環境下でのみ解析が行われた。データの保管ならびに閲覧/解析については、上記ハードディスクまたはスタンドアロンPCのみでのみ行われ、その他の機器等での利用は行われない形式とした。

上述の通り、本事業では個人情報が含まれるデータの解析を行っている。そのため、受託者が所属する組織(産業技術総合研究所)に設置された倫理委員会(人間工学実験委員会)での

個人情報取り扱い研究該当審査、ならびに、外部組織における倫理委員会(社会福祉法人 恩賜財団母子愛育会 愛育研究所 倫理委員会)の承認を取得した。

<倫理審査>

【産業技術総合研究所 人工知能研究センター 人間工学実験委員会】

- ・整理番号: 人 2017-772(新規)
- ・実験課題名: 児童相談所子ども虐待に関わるデータを用いた確率モデリングとリコメンド開発研究
- ・判定結果: 人間工学実験審査申請非該当

【愛育研究所倫理委員会: 研究計画審査結果】

- ・受付番号: 第4号
- ・研究課題: 児童相談所及び市区町村の児童相談データベースを活用した AI 応用研究
- ・判定: 承認(承認番号第4号)

2.2 データ抽出手続き

2.2.1 三重県データベースからのデータ抽出

虐待対応に関するデータは、県に設置された各児童相談所が共通して利用するデータベースから取得した。具体的には、相談台帳、リスクアセスメント情報、一時保護台帳、措置台帳の4つのデータソースから虐待通告に該当するデータを抽出した。相談台帳には、児童相談所で対応のあった虐待通告を含む相談事例の情報が蓄積されている。性別や年齢等の基本情報は通告受理段階で入力され、一時保護の実施や終結に関する情報は、当該イベント発生段階で入力され、更新される。リスクアセスメント情報は、虐待通告時及び現認時のリスクアセスメント情報と対応判断が、それぞれ通告受理、現認時、対応判断実施の段階で記録される。一時保護台帳及び措置台帳には、児童虐待を根拠とした事例を含む、一時保護及び措置に関する情報が記録されている。一時保護及び措置に関する情報は、その実施があった事例のみが記録され、実施及び終結・解除時に情報が記録・更新される。

4つのデータソースに記録されたデータは、2018年2月8日にA県児童相談所のデータベースから抽出された。対象データは、2014年4月1日から2018年11月31日までに通告受理・対応記録が入力完了している虐待通告レコードとした。なお、データ抽出時の段階で、氏名や住所等の個人情報項目は検索対象から除外された。解析は、全てIDによる匿名化された状態で実施された。

データソースレベルにおけるレコード抽出基準は、(1)虐待通告事例、(2)2014年4月1日以降に虐待通告として受理され、2018年11月31日までに対応判断までの入力が完了されたレコードであった。このとき、欠損値や異常値に関する処理は実施していない。

データソースからの検索項目について、相談台帳に対しては、基本情報・基本対応情報に関する115項目を検索の対象とした。リスクアセスメント情報からは、緊急受理会議時及び現認時のアセスメント情報に関する204項目を選択。一時保護台帳からは、保護期間や保護理由に関する9項目、措置台帳からは、措置区分や委託期間に関する9項目を対象とした。

上記項目でデータソースから抽出したレコードを解析に利用する際には、次の5つの基準に該当する項目を除外した。すなわち、(1) データベース構成上の予備項目であり入力データが存在しない項目、(2) データソース間で共通する二つめ以降の重複項目、(3) 他の項目に情報を代表させることができる項目(e.g. 区分情報と詳細区分情報)、(4) 識別子としての機能のみを有する項目、(5) 地域に特有で内容に一般化可能性が認められない項目を除外した。プロジェクトレベルにおけるレコード抽出では、再発をフォローするための追跡期間(後述)を設定した理由から、通告受理日が2017年3月31日までのレコードを組み入れ基準として解析に利用された。

欠損値は、該当フィールドに欠損ラベルを付与することで、「欠損しているという情報」として活用した。異常値に関しては、保護者の年齢において認められた13歳未満のレコードを除外した。

4つのデータソースは、無作為に割り付けられた児童IDと、反復出現する児童IDの出現番号によって結合された。このとき、相談台帳データに出現するレコードに対して、結合可能な一時保護台帳及び措置台帳のレコードを紐づけた。データの加工及び結合処理は、統計解析環境R(version 3.5.0)を用いた処理プログラムによって実行され、処理プロセスはRパッケージrmarkdown(version 1.10)によって記録された。

2.2.2 神奈川県データベースからのデータ抽出

神奈川県の児童相談所が扱う児童虐待対応に関するデータを抽出。データソースレベルにおけるレコード抽出基準は、(1)平成20年4月1日から平成30年3月31日まで、(2)主たる相談種別が「養護・虐待」とした。管轄所内にて関連データを抽出し、同所内にて個人情報の匿名化処理を実施した。抽出したデータは全て閉止済みのレコードであった。抽出した項目は相談台帳ならびに対応に関する情報であり、受付年月日、閉止年月日、児童の性別と年齢、相談経路、閉止時の担当者情報(職種・ID)、処理状態(措置内容)であった。また、児童の状況をもとに判断づけられた1から8(1が最も重篤)の整数値によるリスクアセスメント項目も含まれていた。ただし、リスクアセスメント情報は継時的に更新が加えられる項目のため、閉止時の情報となっていた。欠損値は、該当フィールドに欠損ラベルを付与することで、「欠損している」という情報を与えて活用した。ただし、リスクアセスメント項目が「0」とミス入力されているレコード(24件)は除外した。なお、データソースには住所情報が含まれたため、丁目より詳細な情報を全て除外し、児童IDと担当者氏名は全てID化した。所内における個人情報の匿名化・除外処理の結果は、現場担当職員の確認を経たのちに、解析に利用された。

2.2.3 千葉県データベースからのデータ抽出

虐待対応を含む児童相談所業務に関するデータを抽出。県に設定される児童相談データベースから、相談台帳、一時保護台帳、措置台帳の3つのデータソースから、個人情報を含まないデータ項目が抽出された。相談台帳には、三重県と同様にして、児童の性別や年齢、受付年月日などの基礎情報が格納されており、その他の項目として健診に関する受診の有無や買春等被害相談の有無などが含まれた。一時保護台帳には、一時保護の開始日ならびに委

託日数、保護解除後の措置先種別などが含まれている。措置台帳に関しては、入所日や対処日、措置回数などが含まれた。これらのデータを解析に利用する際には、三重県と同様の5つの基準で項目の選択を行った。これに加えて、該当レコード数が10件未満の項目に関しても、同様に項目から除外した。さらに、虐待以外の相談事例を全て解析対象から除外した。欠損値に関しては、三重県と同様に「欠損している」という情報を付与することで解析に組み込んだ。

これらの3つのデータソースは、無作為に割り当てられた児童IDと反復出現する児童IDの出現番号によって結合された。このとき、相談台帳データに出現するレコードに対して、結合可能な一時保護台帳及び措置台帳のレコードを紐づけた。

2.3 自治体 DB の項目整理

各自治体のデータベース情報をもとに、主要台帳で用いられている項目について整理する。なお、一部公開不可である項目については記載しない。また、データベースシステムに含まれる全ての項目を記載することは量的に困難であるため、虐待対応に直接的に関連する情報項目についてのみ抜粋して記載する。

2.4 虐待再相談と対応期間の予測手法

本稿では、データベース項目が持つ予測の有用性について検証するためのデータ解析を行う。すなわち、データベース項目を活用することで事例の見通しや、それに基づく対応方針への示唆を得ることがどの程度可能かについて、具体的な事例を示す。その際の方法として、(1) 再相談予測に関わる要因の検討、(2) 事例の対応期間の予測、(3) 再相談と対応期間の高精度予測を行う。それぞれの手法について、以下に整理する。

【1】 ロジスティック回帰モデル

再相談予測の関連要因を検討するモデルとして、ロジスティック回帰モデルを利用する。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」についての係数値を得ることができる。これらの値は、「他の説明変数が持つ関連性を調整した上で」の関連度合いを示すため、当該項目単独の関連性の強さと解釈することができる。例えば、「児童の年齢が若いほど再相談の可能性が高い」という結果が得られた場合、「他の変数(例えば保護者の年齢)の影響を調整した上でであっても、当該の関連性が認められた」と解釈することができる。なお、ロジスティック回帰で得られた各項目に対する係数の値は、自然対数の底を用いた指数変換を行うことで、オッズ比と等価になる。つまり、「児童の年齢が1歳上がる前と比べて、1歳上がったとすると再相談になる確率が何倍になるか」という確率の比で解釈することができる。複数の変数を用いたロジスティック回帰を行った場合に得られるオッズ比は「調整オッズ比」と呼ばれ、他の変数の影響を調整した上でのオッズ比を示す。本稿では、ロジスティック回帰の結果得られた調整オッズ比の値から、再相談に関わる要因を報告する。その際、データベースに含まれる項

目数が極めて多いため、オッズ比の95%信頼区間が1を含まなかった項目のみを抜粋して報告する。すなわち、統計的に関連性が認められた要因に絞って報告を行う。

【2】 負の二項分布を用いた回帰モデル

事例の対応日数に関わる要因を検討するモデルとして、負の二項分布を用いた回帰モデルを適用する。事例の対応日数は、0を最小とする整数値である。比較的短期間の場合が多く、稀に長期的な値をとる右に裾の長い分布形状を持つ。このような、右に裾の長く(分散が大きく)かつ非負の整数値をとる変数を予測する際には、負の二項分布が一般的に用いられることが多い。データに即した統計分析モデルを用いることで、可能な限り予測性能の高い結果を得ることができる。負の二項分布を用いた回帰モデルで得られる結果は、項目ごとに算出される係数値となる。値の絶対的な意味を解釈することはできず、相対的な比較となるため、留意されたい。本稿では、係数の95%信頼区間が0を含まなかった項目の結果のみを報告する。すなわち、「○○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」といった関連性に関する結果を得ることができる。

【3】 機械学習

上記のような解釈可能な統計解析手法とは異なり、「予測性能・的中性能」に重点を置いた解析として、複数の機械学習モデルを利用する。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。また、「本当に再相談になった事例の何%を正しく拾い上げられたか: 感度または Recall」や「再相談だと予測された事例のうち、何パーセントが実際に再相談だったか: Precision」、さらに、「再相談ではない事例を、何パーセント正しく予測できたか: 特異度」などの複数の指標で予測性能を報告する。これらを総合的に評価するための1つの指標として、ROC 曲線と曲線下面積 (Area Under the Curve, AUC) という指標も合わせて報告する。これらの指標は全て0から1までの値を取り、1に近いほど高い性能を示す。

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、ベ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。XGBoost のパラメータは、学習率(η) = 0.1、Tree depth = 7、更新回数 = 20000 とした。ランダムフォレストのパラメータは決定木の深さが5、復元生成する決定木の本数が500 とした。また、多層ニューラルネットワークは、中間層を3層とし各層のパラメータを43・38・13 とした。学習率は0.0049、活性化関数はシグモイド関数とし、目的関数は精度とした。なお、このチューニングは初期値を500回投入、更新回数を10とするベイズ最適化により探索した。各機械学習の結果を統合(アンサンブル)する際には、単層ニューラルネットワーク(活性化関数はシグモイド関数)を利用した。

3. 結果

3.1 児童相談所におけるデータベース項目の例

3つの自治体のデータベース仕様を元に、解析的に活用可能性のある項目について抜粋、整理を行った。なお、以下の整理は既存の自治体データベースに含まれている項目のみを対象としているため、必ずしも「これらの項目で十分」であることを意味しない。

【相談台帳項目例】

- 児童番号(あるいはケース番号)
- 受付年月日(あるいは通告年月日)
- 受付時間(あるいは通告時間)
- 受付経路
- 通告一次経路
- 受付区分(新規・対応中再通告・終結後再受付)
- 管轄児童相談所
- 児童権限区分
- 相談所コード
- 担当課コード
- 主たる虐待者
- 従たる虐待者
- 主たる虐待種別
- 主担当者番号(あるいは担当者氏名)
- 心理担当者番号(あるいは心理担当者氏名)
- 児童年齢
- 児童学齢区分
- 児童権限区分
- 児童性別
- 保護者年齢
- 保護者性別
- 家族構成
- 閉止年月日
- 住所
- 障害種別
- 障害程度
- 身体障害者手帳の有無
- 等級
- 処理状態
- 虐待者本人来所
- ぐ犯行為
- 触法行為
- いじめ相談

- 児童買春等被害相談
- DV・面前暴力事案
- 安全確認件数
- 校区
- 親権後見人関係項目

【一時保護台帳項目例】

- 地域
- 一時保護決定日
- 一時保護開始年月日
- 一時保護終了年月日
- 一時保護開始理由
- 一時保護の委託期間
- 保護の種別
- 解除後の措置先種別
- 解除後の措置先名称
- 一時保護先区分(所内・委託)
- 一時保護先種別
- 一時保護先名

・措置台帳項目

- 措置先名
- 種別
- 措置開始日(入所年月日)
- 措置停止日
- 停止解除日
- 退所年月日
- 退所理由
- 措置回数
- 市町村名

【その他業務関連台帳項目例】

- 職員配属年月日(職員リスト)
- 一時保護行動診断表
- 里親台帳
- 進行管理台帳

【リスクアセスメント項目例】(ただし、詳細は非公開)

- 緊急出動検討項目
- 一時保護検討項目
- 初期調査検討項目
- ニーズアセスメント検討項目

3.2 基礎集計

本節では、各自治体のデータについて、基礎集計を行う。ここでは、解析に利用する再相談率と対応日数、そしてデータサイズである通告件数の集計のみを行う。なお、ここでの「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。ここで、自治体によってデータベースへのレコード登録過程が異なることに留意されたい。例えば、一度の通告から終結までの間で再度の通告があった場合に、それを再度データベースにレコードとして記録するか否かなどで、自治体に差がある場合がある。そのため、同じ「再相談」であっても、自治体によってその質的な意味合いが異なっている。したがって、対応期間についても同様に、自治体間でその値の大小を比較できないことには注意が必要となる。

3.2.1 三重県の基礎集計

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 2022 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(図 2-1)。

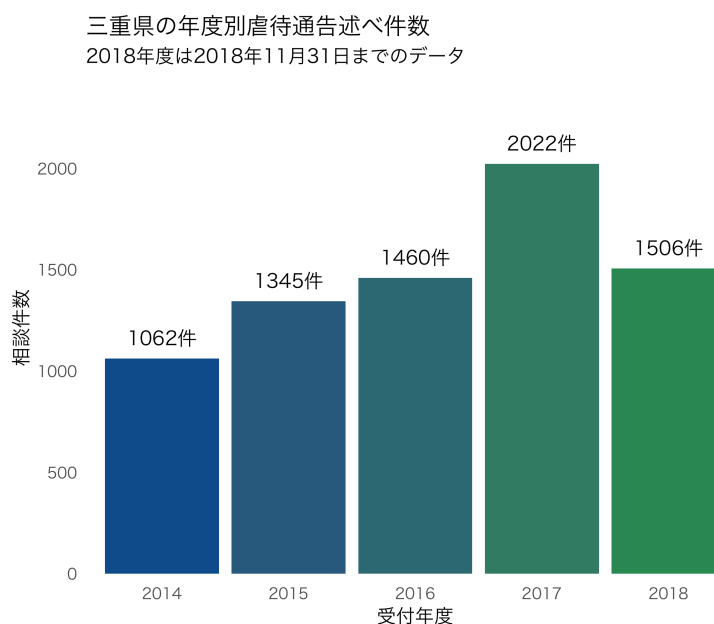


図 2-1 三重県: 年度別虐待相談対応件数(述べ)

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(図 2-2)。

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)

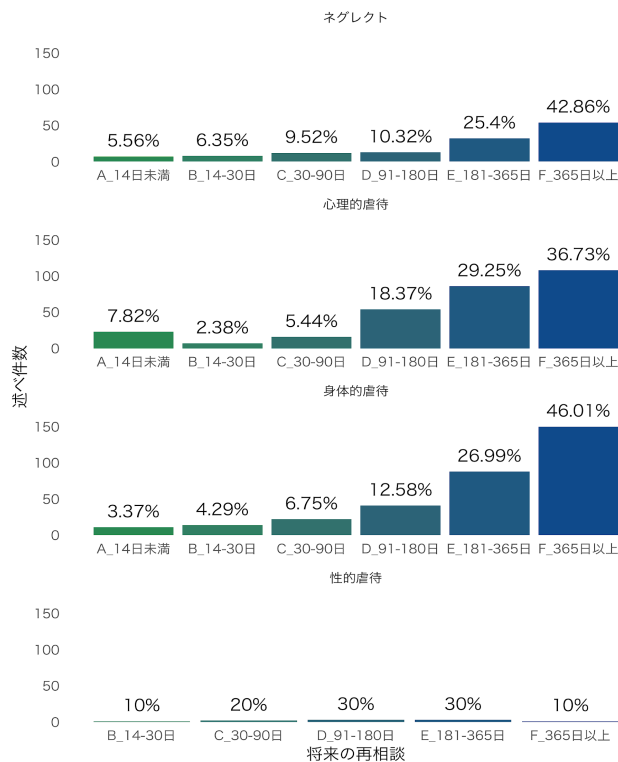


図 2-2 三重県: 虐待種別 対応期間区分の該当割合

また、延べ件数で計上した再相談の割合は、25.41%であり、全体の4分の1程度を占めた(図 2-3)。

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)

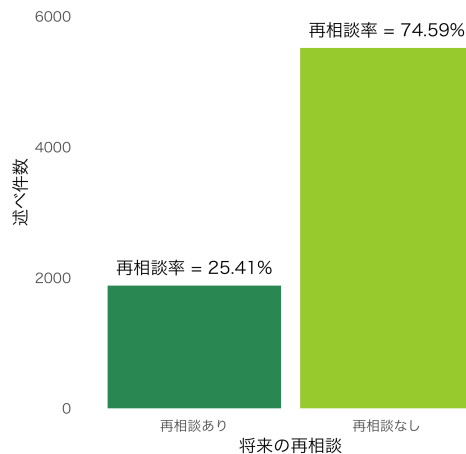


図 2-3 三重県: 延べ件数を利用した再相談件数の割合

3.2.2 神奈川県基礎集計

神奈川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 24590 件の虐待相談記録であった。年度別の虐待通告件数は、2016 年度で 3222 件となっている。2018 年度については、3 月 31 日時点までに終結となっていないデータが含まれているため、参考値となる(図 2-4)。

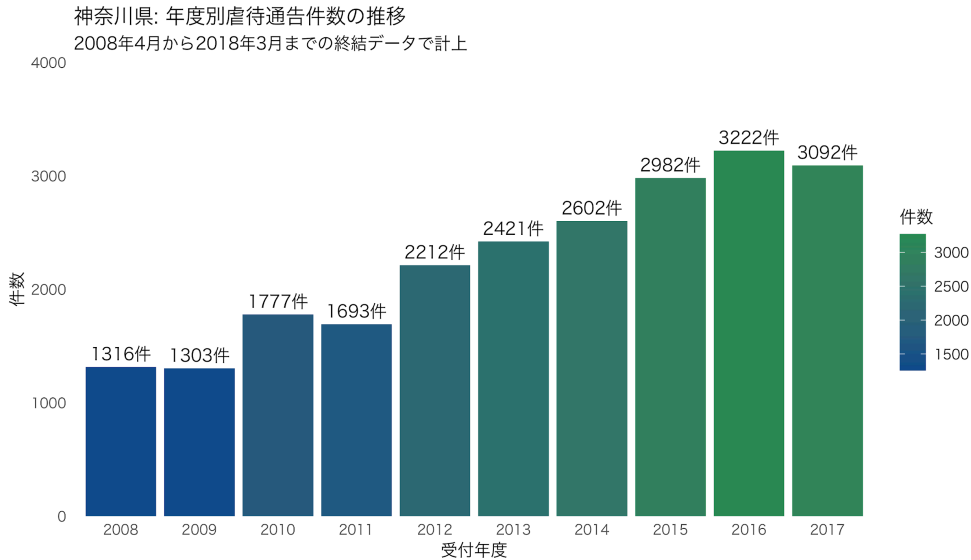


図 2-4 神奈川県: 年度別虐待相談対応件数(延べ)

種別の対応日数では、第一に 14 日未満で終結する事例群が存在し、30 日以上対応となる群の 2 つが存在しているように観察される。通告件数の多い心理的虐待事例では、事例の 4 分の 1 程度が 91 日から 180 日程度の対応日数となっていた(図 2-5)。

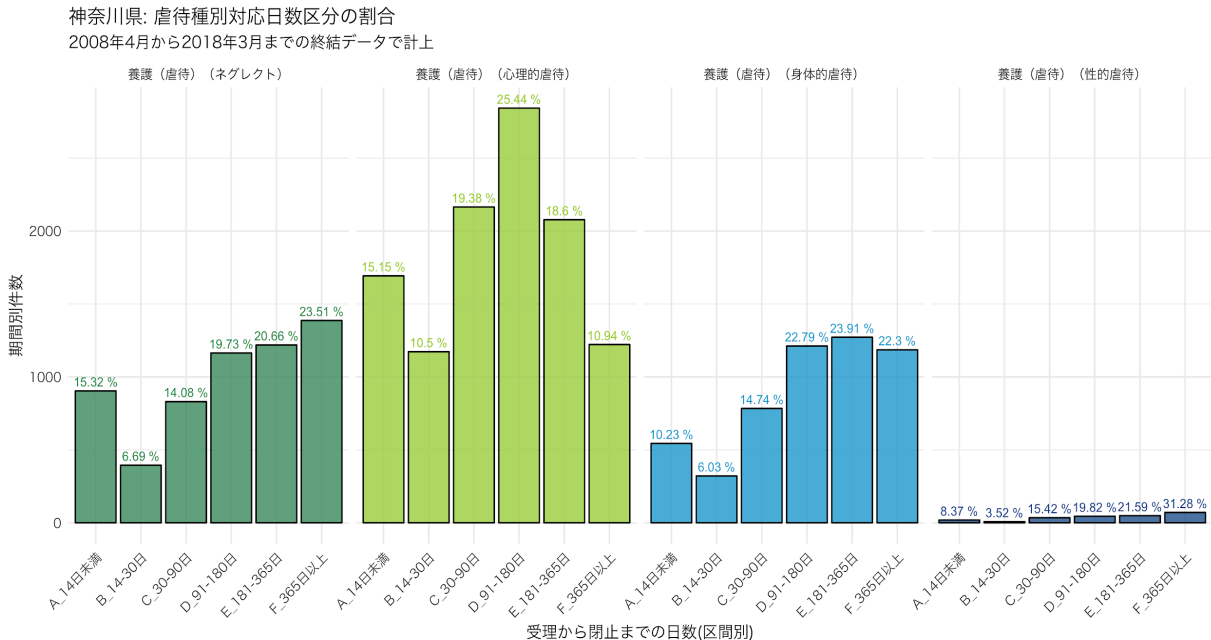


図 2-5 神奈川県: 虐待種別 対応期間区分の該当割合

虐待相談の再相談率については、延べ件数を用いてカウントしたところ、全体の 18.76% と

なった。神奈川県の場合、虐待相談対応のおよそ5分の1程度は再相談によるものであることが示唆された。

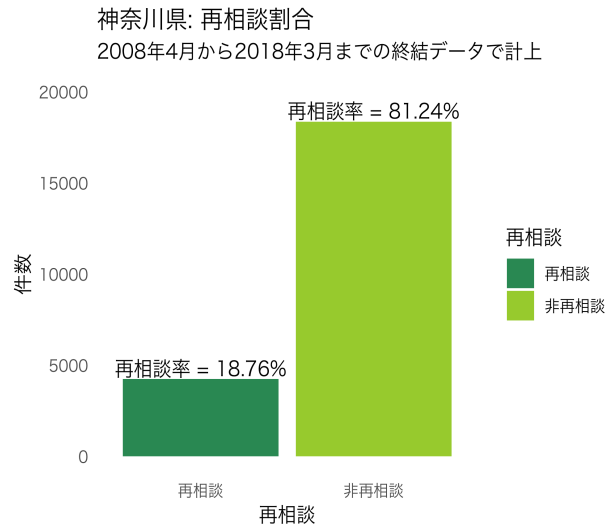


図 2-6 神奈川県: 延べ件数を利用した再相談件数の割合

3.2.3 千葉県の基礎集計

千葉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 30212 件の虐待相談記録であった。年度別の虐待通告件数は、2016 年度で 6522 件となっている。2017 年度については、12 月 31 日時点までのデータであるため、参考値となる(図 2-7)。

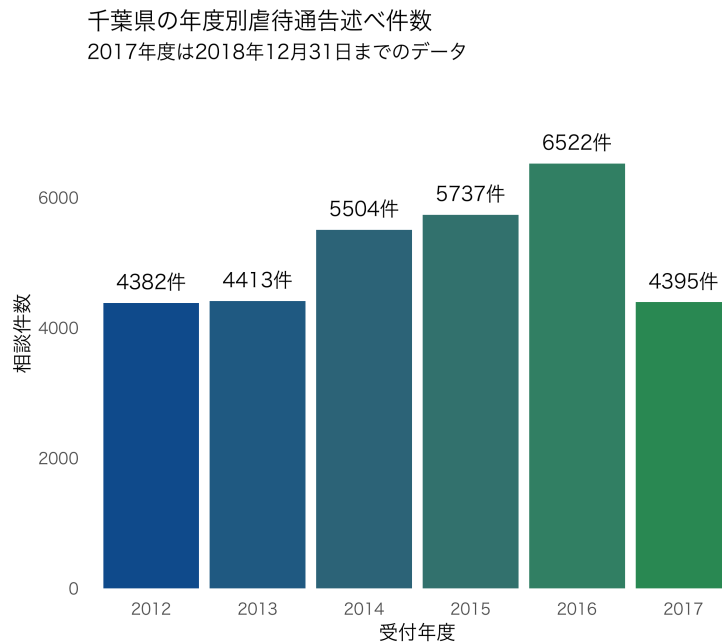


図 2-7 千葉県: 年度別虐待相談対応件数(延べ)

種別の対応日数については、30212 件全てについて算出することができ、その結果、過半数の事例が 90 日未満での終結となっていた(図 2-8)。

千葉県: 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)

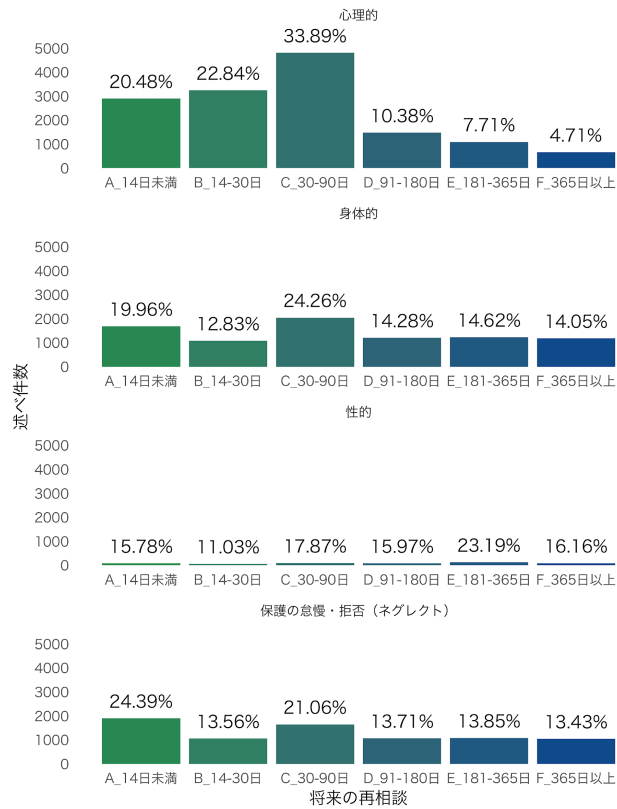


図 2-8 千葉県: 虐待種別対応期間区分の該当割合

虐待相談の再相談率については、延べ件数を用いてカウントしたところ、全体の 37.4% と なった。虐待相談のおよそ 3 分の 1 程度は、再相談によるものであることが示唆された。

千葉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)

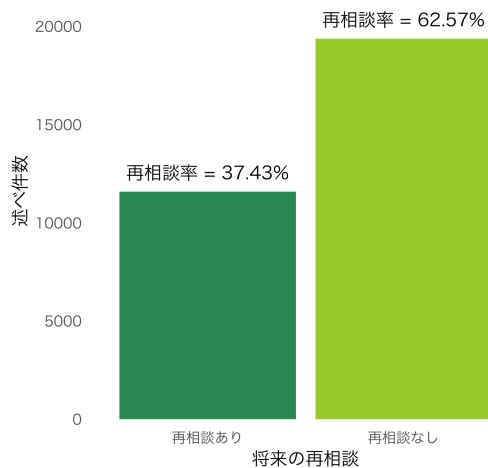


図 2-9 千葉県: 延べ件数を利用した虐待再相談の割合

3.3 ロジスティック回帰を用いた再相談予測と関連要因の検討

3.3.1 三重県における再相談予測モデルの結果

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(ただし、年齢が若いほど: 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 = 1.173, 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(図2-10)。

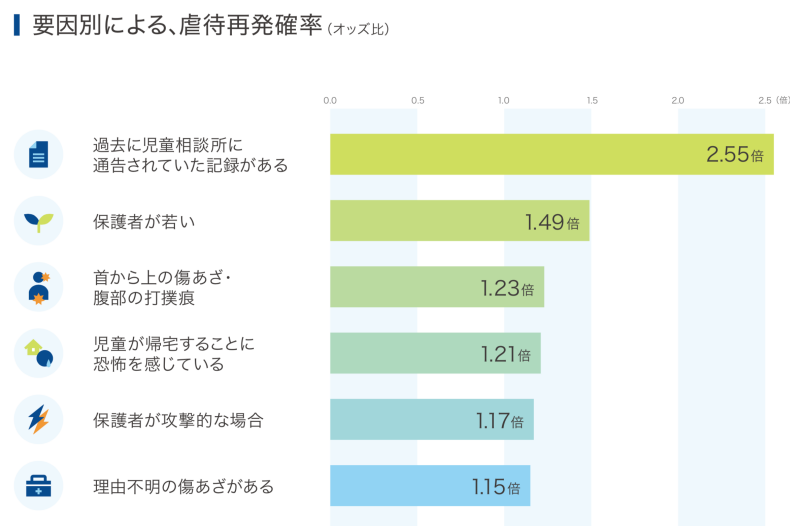


図2-10 三重県：将来の再相談に関わる関連要因(調整オッズ比)

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。

3.3.2 神奈川県における再相談予測ロジスティック回帰の結果

神奈川県のデータに対して、将来の再相談の有無を目的変数とするロジスティック回帰を実施した。推定の結果得られた調整オッズ比を算出したところ、積極的に将来の再相談と関連したのは、障害等級4級(OR = 2.95)、障害程度 = 4(OR = 1.77)、障害程度 = 1(OR = 1.65)、保育所からの通告・相談(OR = 1.62)、その他の相談経路(OR = 1.49)、親戚からの相談・通告(OR = 1.46)、父親からの相談・通告(OR = 1.22)、児童の年齢が幼いほど(OR = 1.05)であった(図2-11)。

神奈川県: 再相談に対するロジスティック回帰
担当職員の影響を調整した調整オッズ比

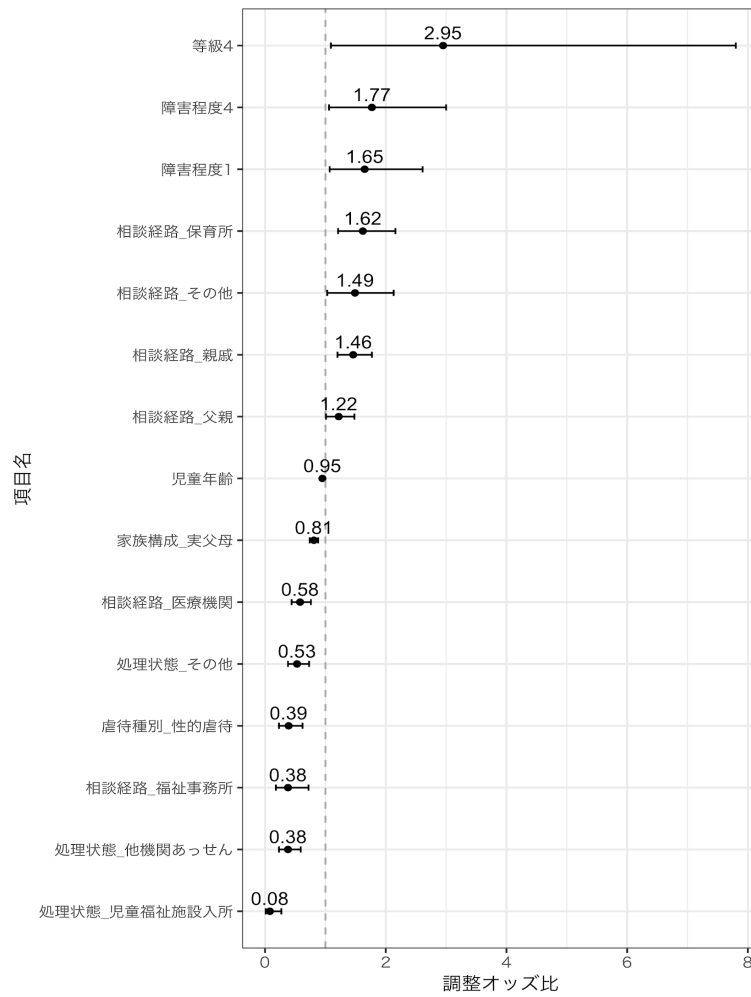


図 2-11 神奈川県: 将来の再相談に関わる関連要因(調整オッズ比)

3.3.3 千葉県における再相談予測ロジスティック回帰の結果

千葉県における、3つの相談台帳を結合したデータを利用したロジスティック回帰モデルによる再相談の予測を行った。その結果、措置の実施(OR = 5.15)、過去の再相談歴(OR = 2.49)、児童福祉施設からの通告(OR = 2.3)、主たる虐待者が不明な場合(OR = 1.52)、母親からの相談(OR = 1.33)、一時保護の実施(OR = 1.28)、親戚からの相談(OR = 1.27)、実父以外の父親からの相談(OR = 1.24)、学校からの通告(OR = 1.21)、男児の場合(OR = 1.16)、児童の年齢が若い場合(OR = 1.02)において、再相談の可能性が高まることが示唆された(図 2-12)。

千葉県: 再相談に対するロジスティック回帰
担当職員の影響を調整した調整オッズ比

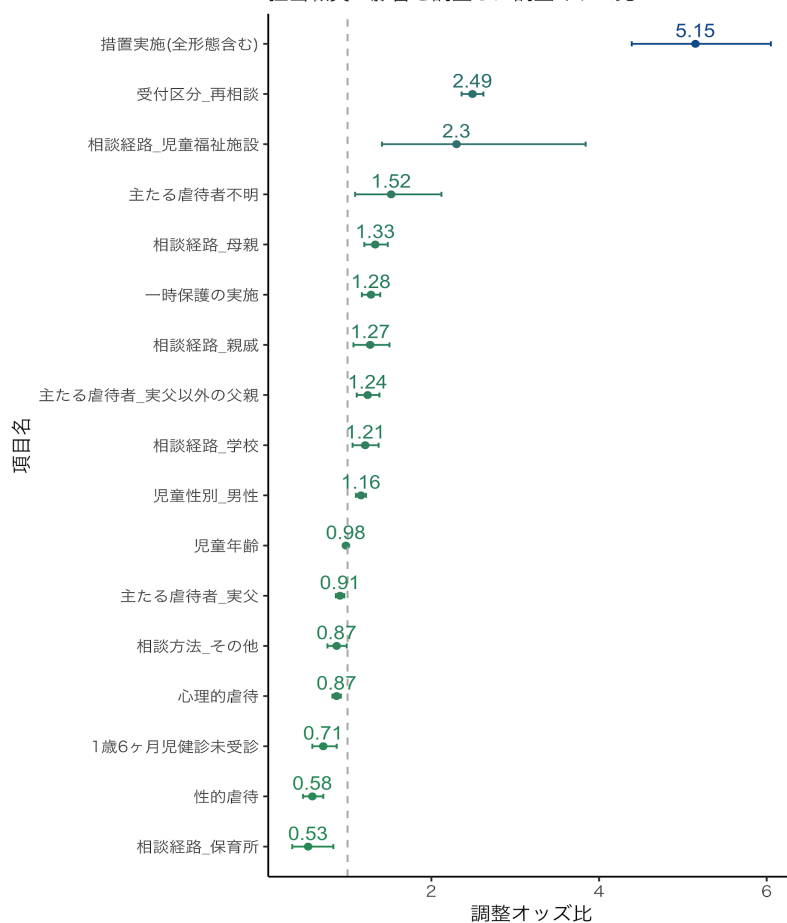


図 2-12 千葉県: 将来の再相談に関わる関連要因(調整オッズ比)

3.4 対応期間予測と長期化要因の検討

3.4.1 三重県: 事例の対応長期化関連要因の検討

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(図 2-13)。

Ⅰ 虐待事例対応が長期化する要因



図 2-13 三重県: 事例の長期化に関わる要因

3.4.2 神奈川県: 事例の対応長期化関連要因の検討

神奈川県データを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。

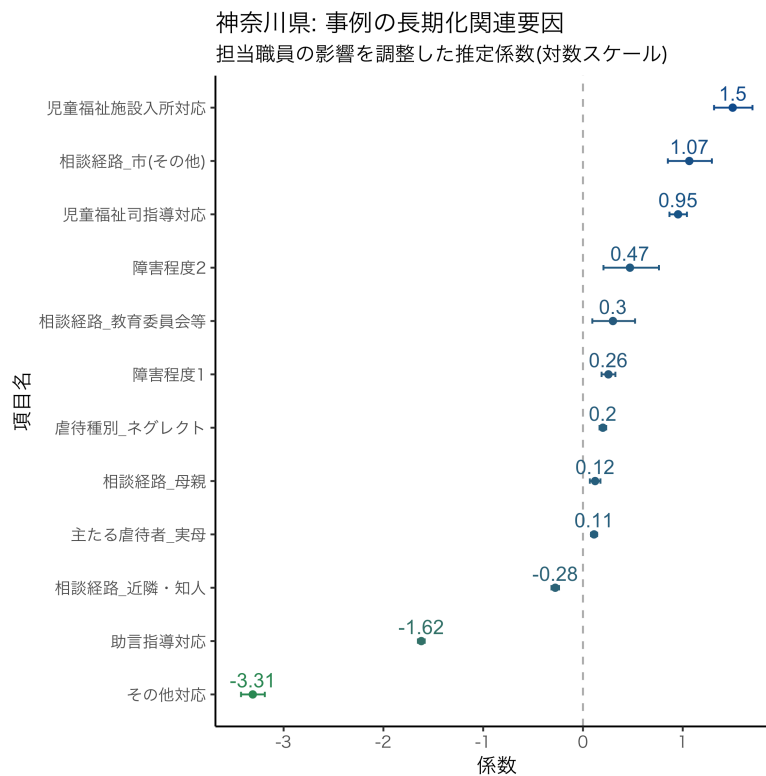


図 2-14 神奈川県: 事例の長期化に関わる要因

3.4.3 千葉県：事例の対応長期化関連要因の検討

対応期間の長期化と関連する要因について検討した。その結果、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(図 2-15)。

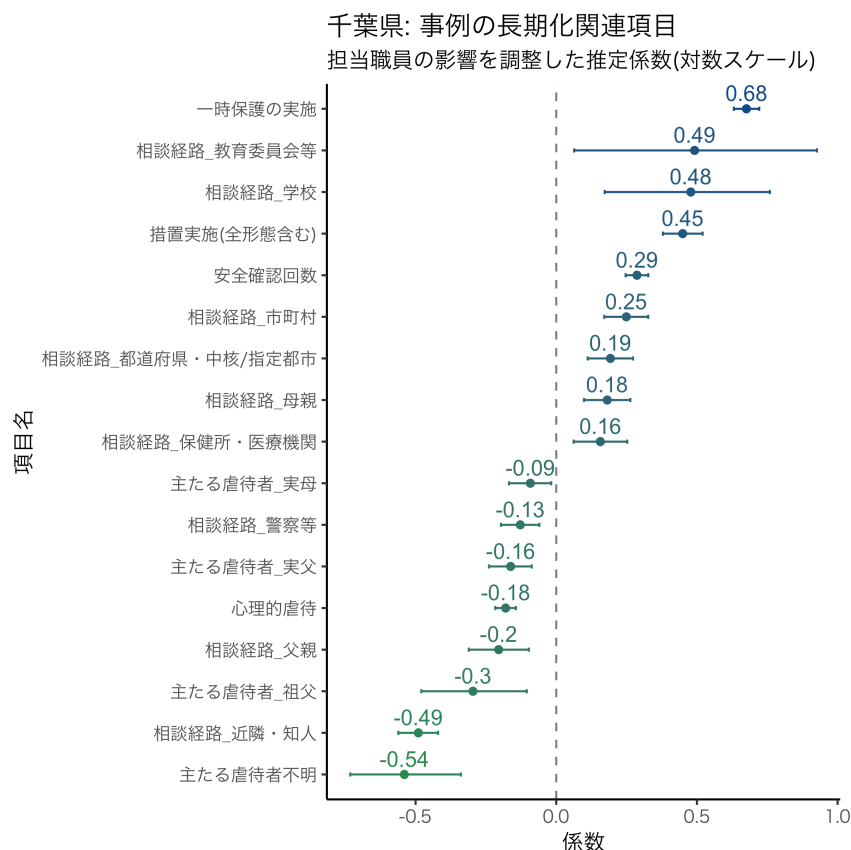


図 2-15 千葉県：事例の長期化に関わる要因

3.5 機械学習による再相談事例と長期化事例の予測

各自治体のデータを活用し、将来の再相談と対応日数を予測する機械学習モデルを構築した。目的は、データベースに含まれる情報量を確認することである。機械学習の精度を以下に示す。結果の詳細は、以降の解析結果を参照されたい。

表 2-1 各自治体データを用いた機械学習の予測精度(Accuracy)

	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%

3.5.1 三重県における再相談予測モデルの構成

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(巻末資料 2)。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。

3.5.2 神奈川県における再相談予測モデルの構成

神奈川県のデータでは、XGBoost を用いた再相談の予測モデル構築を行った。学習データは無作為に選抜した 80%のデータを利用し、残りの 20%を検証用データとした。その結果、再相談を予測する精度は 0.824 となり、実際に再相談となっている事例を再相談であると予測可能だった割合は 42.2%となった。具体的には、再相談ありと予測して実際に再相談だった件数が 174 件、再相談なしと予測して実際に再相談ではなかった事例が 1689 件となった(図 2-16)。他の詳細な結果は巻末資料 3 に掲載している。

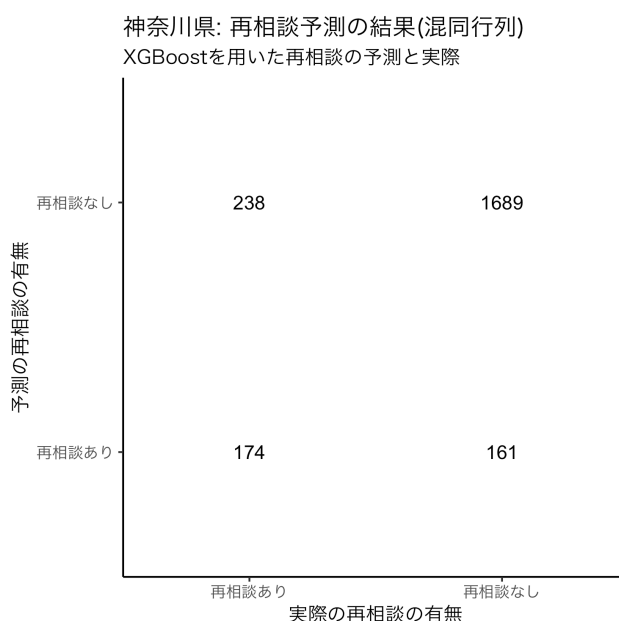


図 2-16 神奈川県: 再相談予測の結果

3.5.3 千葉県における再相談予測モデルの構成

千葉県のデータにおいても、一般的に高い精度が得られるアルゴリズムである XGboost を用いた再相談の予測モデル構築を行った。学習データは無作為に選抜した 80%のデータを利用し、残りの 20%を検証用データとした。その結果、再相談を予測する精度は 0.746 となり、実際に再相談となっている事例を再相談であると予測可能だった割合は 58.5%となった。具体的には、再相談ありと予測して実際に再相談だった件数が 656 件、再相談なしと予測して実際に再相談ではなかった事例が 1652 件となった(図 2-17)。他の詳細な結果は巻末資料 4 に掲載している。

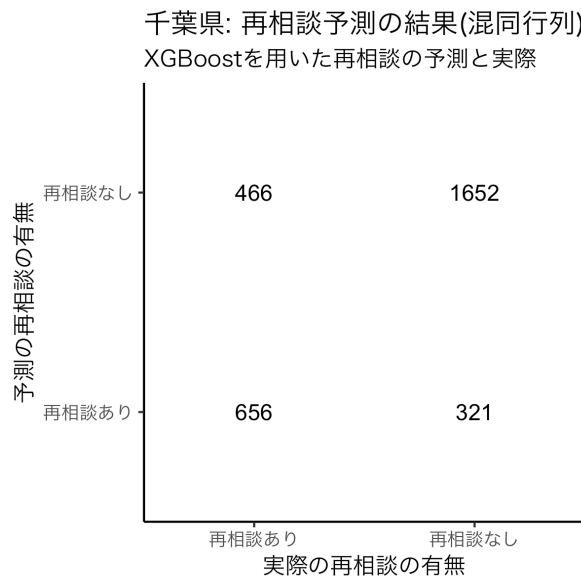


図 2-17 再相談予測の結果

3.5.4 三重県: XGBoost による対応期間の予測モデル

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(多クラス分類)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(巻末資料 5)。

3.5.5 神奈川県: XGBoost による対応期間の予測モデル

神奈川県データベースを用いた機械学習モデルによる対応日数の予測を実施。ランダムに選抜したデータの 80%を学習データとし、残りの 20%を検証データとした。目的変数には対応日数の実数をそのまま利用し、得られた予測日数と合わせて 6 カテゴリに分割。得られたカテゴリでの多クラス分類精度を評価対象とした。その結果、予測精度は 0.375 となった(巻末資料 6)。

3.5.6 千葉県: XGBoost による対応期間の予測モデル

千葉県データベースを用いた機械学習モデルによる対応日数の予測を実施。ランダムに選抜したデータの 80%を学習データとし、残りの 20%を検証データとした。目的変数には対応日数の実数をそのまま利用し、得られた予測日数と合わせて 6 カテゴリに分割。得られたカテゴリでの多クラス分類精度を評価対象とした。その結果、予測精度は 0.422 となった(巻末資料 7)。

4. 考察

4.1 三重県の結果に対する考察

三重県のデータベースを用いた解析の結果、将来の再相談の関連要因、対応日数の関連要因、再相談と対応日数の予測性能(データが有する情報量)が示された。例えば、「首から上に傷あざがある・腹部の打撲痕がある場合」などは、将来に再度の通告が生じる可能性が高まるといった結果が得られた。このような、現場で目にする多くの児童や家族の状態像にそぐう項目によってその予後を予測できることの有効性は大きい。それは過去の実例から導かれた「知識」であり、当該状況に遭遇した際の見通しを持つことが可能となる。将来の再相談可能性や、事例の長期化(危険な環境下で児童が過ごす時間の長さ)に関する見通しを得ることで、対応の方針をより慎重に検討することが可能となる。

ここで、再相談予測に関して具体的に得られた知見について考察する。各自治体に共通する部分があるが、過去の係属歴がある場合には、さらなる将来の再相談が発生する可能性が高いことが明らかとなった。反復する事例(児童が継続的に危機に晒される可能性のある事例)への対処にあたっては、強制的な介入措置(一時保護等)を実施することが場合によっては有効である可能性が示唆された。保護者の年齢が若いほど再発可能性が高いという結果については、通告受理段階から今後の継続的な虐待の可能性に留意することが可能な知識である。また、児童が帰宅することに恐怖を感じている場合、保護者が攻撃的な態度を有する場合においても、将来の再相談が懸念された。在宅措置をとることによって、虐待が継続され、後の再相談につながっていることが推測される。このような場合においても、一時保護を含む主たる虐待者との物理的な分離措置を講じることが後の虐待継続を防止する可能性が指摘できる。

次に、対応日数の長期化要因に関しては、複数の利用可能性がある。第一に、対応現場職員のマネジメントに関するメリットが得られる。あらかじめ長期化が見込まれる事例については、その負担を考慮した担当職員への割り振りを実施することができる。職員負担を考慮したマネジメントに繋がるだろう。また、事例の長期化があらかじめ見込まれることで、担当するケースワーカーは対応上の見通しを得ることができる。緊急的な対応が必要な場面と、市区町村への見守りへの切り替えなど、長期的対応を前提とした支援方針の策定を事前に講じることにつながる可能性がある。

機械学習による再相談と対応日数予測では、高い予測性能を得ることが可能となった。これは、リスクアセスメントを含む三重県データベースに「予後の予測を行えるだけの情報が含まれている」ことを意味している。ただし、約7000件のデータであるため、確定的なパターンを発見することは難しく、精度に一定程度の揺らぎが生じていることには留意されたい。過去のパターンに基づく機械学習の知見は、各職員が経験可能な事例の範囲を超えた情報として、有効に活用できるだろう。なお、機械学習の予測結果は、「チューニング」により使い方を調整することが可能となる。「再相談となりうる事例を可能な限り取りこぼさずに拾いたい」といった場合や、「再相談の可能性が極めて低い事例を検出し、早期の終結対応につなげたい」といった、虐待継続の見逃しリスクの軽減や、業務負担の軽減に活用することが可能となる。

4.2 神奈川県の結果に対する考察

神奈川県における再相談の割合は全体のおよそ 18%であった。職員による事例対応業務のおよそ 5 分の 1 が、反復する通告に対する対応であると解釈できる。中でも、通告件数の多い心理的虐待については、その 4 分の 1 程度が 91 日以上に対応事例となっていた。再相談の可能性や、長期的対応についてあらかじめ予測することが可能となれば、事前の対処や長期化に向けた効果的な対応を検討することにもつながるだろう。

再相談の関連要因に関する解析では、並存する身体障害者手帳・療育手帳の等級、通告経路に関する情報が抽出された。他の相談種業務との並存による再相談記録の増加傾向や、通告経路によって事例の重篤性や特徴が異なることが推測される。

障害等級などは、この事例対応時点での一つの推測指標となりうる。しかし、事例の内容的側面に関する情報不足のために、「どのようなケースだと、再相談や長期化につながるのか」といった現場で活用可能な事例の特徴把握には至らなかったと結論できる。対応期間の長期化関連要因についても、同様の結論を得る結果となった。すなわち、障害等級や通告経路を主とする外堀の情報が事例対応の長期化を予測した。事例の内容が通告経路によって大きく異なり、それゆえに対応期間にも差が生じているものと推測されるが、それ以上の解釈を得ることは難しいという結果となった。

再相談と対応日数を予測対象とする機械学習の結果は、その性能に関して十分な結果を得ることができなかった。再相談と対応日数という側面から鑑みて、データベースに記録されたデータに十分な情報量が含まれていないことを示唆している。年々増加する通告件数に対しては、「必要な支援を必要な事例に的確に届ける」ことによる再相談の防止や、対応期間の(質を担保した上での)短期化/効率化が求められる。今後、児童の状態や養育環境の状態に関する情報がデータベースに入力・蓄積されることによって、より高い精度での予後予測など、情報の有効な利活用につながることが期待される。

4.3 千葉県の結果に対する考察

千葉県における再相談の割合は、全体のおよそ 34%であった。すなわち、職員対応業務のおよそ 3 分の 1 が、こういった反復する事例に対する対応となっている。このような再発事例をあらかじめ予測できれば、当該特徴を有する事例に対する重点的なケア対応を行うことが可能となる。効果的な再発防止対応につながれば、児童の安全確保と将来的な業務負担の軽減が期待される。

千葉県のデータベースを用いた再相談に関する解析では、通告経路を中心とする関連要因が抽出される結果となった。通告経路によって事例の重篤性や特徴が異なることが推測される。しかし、児童の状態や養育者に関する情報項目がデータベース上に記録されていないため、「具体的な事例イメージに対応する知識・予測結果」を生み出すことが叶わなかった。このような状態は、再相談に関わるパターンが外堀の情報(通告経路や主たる虐待者の性別など)によって導かれた場合であっても、それを具体的な現場対応に活用しにくいといった側面につながりうる。再相談と対応日数を予測対象とする機械学習の結果は、その性能に関して十分な結果を得ることが叶わなかった。再相談と対応日数という側面から鑑みて、データベースに記録されたデータに十分な情報量が含まれていないことを示唆している。これま

でのケース記録を整然化し、解析的に利用可能な形式でデータベースに蓄積することや、児童の状態・生活環境の情報をリスクアセスメントの一環として項目記録することで、より高い精度での予後予測が実現可能であると考えられる。通常業務の枠組みで支援を充実させるアセスメントと、その記録データを活用することで得られる知見で、より良い対応へとつながることが期待される。

4.4 リスクアセスメントの活用に関して

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、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、事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の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、より客観的な評定方略が定立可能となる。

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、その予後予測に関する知見や重篤度など、これら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が期待される。

4.5 データベース設計への示唆

3つの自治体データ項目について調査した結果、児童相談所において利用されるデータの構造は、基本的部分についてはその多くが共通していた。言い換えれば、「記録」を前提とした項目(受付日時、虐待種別など)については、名称こそ異なるものの、機能的には同一の項目が多数存在した。例えば、厚生労働省が発布する福祉行政報告例に合致する項目などは、多くの自治体のデータベースに組み込みがなされている場合が多いものと推測される。

自治体によって異なる点は、例えば家族形態の記録方法や、特定妊婦、DV・面前暴力事案などの記録項目の有無、そしてリスクアセスメントに関する項目で差異が認められた。ま

た、項目情報の更新方法にも違いが認められた。例えば、事例への接触がなされるたびにリスクアセスメント項目の更新がなされ、常に最新の情報が上書きされる場合と、緊急受理会議時・安全確認時または保護判断時のアセスメント状況が記録される場合などである。このように、同じ項目であっても内部の情報の意味が異なる場合、情報の活用段階での項目の概念的妥当性が担保されないこととなる。標準データベースを構想する上では、入力情報の保持方法についても、統一する必要があるだろう。

そして、リスクアセスメント情報の記録で今後の大きな活用可能性が期待される。事例対応の現場で運用可能な範囲で、可能な限り多角的な視点が記録可能なアセスメント項目の設計が求められると言える。

総合考察

本調査事業について公募要領に記載されていた目的は、次の2点であった。

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すること。その情報集約システムを活用して児童虐待の端緒となる要因を把握し、児童虐待の発生予防・早期発見に資することを明らかにすること。

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。

これに対応する成果物として、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、上記内容を整理・分析すること、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があげられた。

情報連携と関係機関の連携に関する項目①については、その好事例として、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。また、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することで、対応日数や虐待再相談率の関連要因が示された。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、神奈川県及び千葉県のデータベースからも有益な知見を得ることができた。なお、神奈川県、ならびに千葉県のデータベースに組み入れられているリスクアセスメントに関する項目は下位項目を総合した合計値のみとなっていたため、詳細な情報が得られず予測精度が安定しないことが示された。

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の自動出力の有用性を指摘した。当該データベースでは、近年の虐待死亡事例に対する問題点を含め、現場での情報共有が主眼とされる。現場での利便性を基礎として、データに基づく検証が実現できる仕組みとして、その実装が期待される。

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を提言した。

以下では、さらに第一部、第二部で明らかになったことを中心に総合考察を行う。

第一部について

国際比較研究用のデータベースを作るには、現場負担の軽減とデータに基づく検証(研究)の両軸を満たす設計が必要になることを指摘した。情報共有 DB が研究用 DB を兼ね備えることを前提とする設計が重要となる。特に、情報共有 DB は、児童相談所職員が使うため、UI/UX などについても整備された運用が望ましい。そして、2018 年目黒区、2019 年野田市の虐待死亡事例においても問題となった転入・転出の情報共有の漏れを無くすため、自治体間を越えた情報共有を前提にする DB であることが望まれる。この時、情報共有 DB・研究用 DB の構造が標準化/共通化されていることが必須となる。連携の際はもちろんのこと、データを解析する際には統一された項目情報が必要になるためである。

現時点の全国の児童相談所の DB はパッチワーク型で拡張を繰り返した結果、複雑な構造かつその保守運用に多くの費用が生ずる。アメリカ、カナダなどの自治体において、虐待 DB に関する制御は国が行ってきた。個別の自治体間で標準化を検討する際の調整負担が主たる背景要因だと考えられる。実質的な運用を素早く開始するためには、国として標準的な DB 構造を指定することが必要となる。

情報共有 DB の設計に不可欠な点について、ここで改めて 5 つに整理する。

第一に、標準 DB として、リスクアセスメント各項目を含めたデータ構造を標準化することである。どのような項目を入れるのかについては、情報共有用 DB でも、研究用 DB でも、リスクを同じ物差しで管理できるよう、リスクアセスメント各項目を含めたデータが必要となる。これができれば、担当者や所属機関によって同じケースに対してリスクの見立てが違っているとしたら、どういう点でそのような違いが生じているのか、人材育成の点でも検討が可能となる。

第二に、使いにくい DB だと誰もデータを入力しないため、現場職員がデータ入力をしやすいよう、UI/UX を最大限配慮する。2017 年現在、モバイル端末全体(携帯電話・PHS 及びスマートフォン)の保有率は日本国民の 84.0%であるうち、スマートフォンの保有率が 60.9%(前年差 4.1 ポイント上昇)となっている(携帯 4 台に 3 台はスマホ)¹¹。そのため、ほとんどの現場職員が使っているスマホレベルまでには、UI/UX を最大限配慮すべきといえる。それができれば、取扱説明書を読まずにデータ入力や閲覧方法を直観的に理解でき、データが集まりやすい環境を醸成できる。

第三に、情報共有は都道府県をまたいでの情報共有ができるよう、自治体内だけでの情報共有に閉じず、①の共通のデータ構造を活用し、当初から LGWAN にて DB 実装を目指すべきである。最初から自治体内に閉じてしまうと、データ構造が自治体内のベンダーの DB 構造に依存してしまい、標準化からは遠くなる。そのため、標準化及び情報共有をセットで考えることで、導入コストを最小化し、実務にも研究にも使えるベネフィットをあらかじめ最大化するような拡張性を講じることが重要である。

第四に、標準の情報共有 DB ができれば、API などによる中間サーバー制御により、情報共有 DB から研究用 DB へのデータエクスポートを自動化、かつ、個人情報該当項目も自動で削除できる。それができれば、省庁や研究者にとって常に最新のデータを閲覧・取得だけでなく、今後の調査などがあっても研究用 DB から得られないところのみ、現場にアンケート調査協力を依頼するなど、現場の負担を最小限にすることができる。特に情報共有

¹¹ 総務省 | 平成 30 年版 情報通信白書 | 情報通信機器の保有状況

<http://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h30/html/nd252110.html>

DB から研究用 DB への自動出力部分を、プログラム処理等の再現可能な方式で実現できるのであれば、現場アンケートでの入力ミス等も回避されるだろう。

一方、第五に、国際比較のためには、透明性が高くコードブックが開示されている NCANDS の子ども用データベースのデータ構造に近づけることが望ましい。国際虐待防止学会における ICAST なども参照されうる。しかし、コードブックがオープンになっており、透明性が高いもの、かつデモグラフィック要因とリスク関連要因が明らかになっている項目としては、NCANDS が集める子ども情報に関するデータ内容は 2019 年時点では比較的網羅的であると考えられる。ただし、NCANDS が集める組織毎のデータ内容は、コールセンターやディファレンシャルレスポンス、虐待のスクリーン制度や Substantiated / Unsubstantiated の棲み分け基準など、データの記録が発生する施設や体制上の枠組みで本邦と異なる点が指摘される。すなわち、それらの仕組みが整っていない本邦の実態とは異なる体制上で考案された項目であるため、現時点で完全な国際比較を行うためには、データベースの設計そのものに加え、スクリーニングシステムの整備など体制上の課題解決を同時に進めていく必要がある。

業務記録がそのまま統計カウントになるような、情報の記録・共有と利活用が並存するデータベースを設計する。このような取り組みは、業務の効率化だけでなく、児童相談対応の重要なインフラとして活用されうる。なぜならば、虐待対応現場では迅速かつ多層な情報(データ)こそ、現場判断で重要視されるからである。現場で使えるデータが蓄積されれば、上記で示したように、研究用 DB の活用だけでなく、今後 AI によるリアルタイム分析にも活路が見いだせる。他にも、人員配置や予算請求の根拠などにも適用が可能となってくるだろう。

そのため、標準 DB を構築する際には、国際比較だけの目的でなく、社会的養育ビジョンに沿い、現場が子どもを転出入があっても追跡でき、常に最新の情報が入手できるという恩恵を想定した設計が望まれる。また、様々な関連機関との情報共有をあらかじめ見込んでおかなければならない。情報共有 DB の利便性が高まることで、研究用 DB としての価値も相乗的に高まってゆく。

第二部について

標準データベースが整備された環境下では、データの蓄積に伴った複数の活用可能性が期待される。特に、統一されたリスクアセスメント項目の運用とそのデータ蓄積がもつ意義は大きい。

第一に、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る。これらは、現場における業務の均てん化や職員の公平なケースの割り当てなど、データに基づく業務管理に活用することができる。

第二に、事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。もし死亡事例が起こったとしても、どのようなリスクがどの時点で見落とされていたのかを明らかにし、そこからデータに即した再発防止策を考えることも可能になる。

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。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。これらは、人材育成としても重要な知見であり、ケースバイケースと現場で言われてきた言葉も、データに基づく十数パターンの組み合わせで対応している可能性もある。そうであるならば、初任者には優先的によくあるケースパターンから人材研修を行うことであったり、中堅者であれば担当者毎の得意・不得意分野を知った上で、自分でどのような研修を受けたいかを検討し、キャリアプランに活かせることなどが考えられる。

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度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、円滑な連携や知識の集積、効果的な事例対応につながることを期待される。

まとめ

リスクアセスメントの各項目がある三重県と、各項目がない神奈川県、千葉県について、再相談と対応日数の予測を行なった結果、リスクアセスメントの各項目がない場合、ケースの特徴を明らかにする情報量が不足し、機械学習の予測精度が落ちることが明らかとなった。そのため、リスクアセスメントは、合計得点だけでなく、全項目のデータをDB上に保存することが望ましい。また、リスクがどのように遷移したかを評価するために、リスクアセスメントのデータは最終的なデータが上書きされるのではなく、入力タイミングにおいて、全てリスクの遷移情報をデータとして残すことが重要である。

また、データベースは情報共有としての使いやすさをまず網羅することが重要であると繰り返し述べた。特に昨今の虐待死亡事例に対して転入・転出の情報の抜けを無くすことが重要であることから、情報共有をどのように効率化していくのかという点を十分吟味することが、国として虐待再発防止に取り組むアクションになると考えられる。そのため、情報共有は、少なくとも児童相談所では、標準DB構造に合わせて全国の児童相談所が情報共有を開始できるような体制が必要と考えられる。そのためには、自治体向けにセキュリティ要件を満たすネットワークとしてサービス提供しているLGWANを利用し、最新の虐待情報をフィードバックできるようなシステムを根幹から構築することが重要といえる。

最後に、データベースは実務と研究だけでなく、人材育成でも活用が期待される。つまり、研修や実務のマネジメントのためにデータが活用される可能性がある。データ入力に直結するリスクアセスメントは、全国統一的なカウントのルールであったり、マネジメントが前提のため、常に継続的にデータに基づく研修体制をセットで行うことが必要である。

また、情報システムに関する議論では、その開発と実装だけでなく、運用の側面を見落としてはならない。児童虐待対応における情報システムを支える上では、現場のニーズと技術的シーズ、そして、データベース構造を共に理解する人材が必要不可欠である。今後は、弁護士や医師の児童相談所常勤雇用や配置を議論するのと同列に、データベースエンジニアあるいはソーシャルテクノロジーオフィサーの雇用なども検討する必要がある。**忙しい環境だからこそ、全て「人を増やす、人の質を上げる」ことだけに着眼した解決策はその効果**

に不足が生じうる。 Digital Transformation の流れを踏まえ、何を人がやるのか、何を人はやらないのかの業務フロー全体の見直しが必要である。データは、研究者のためではなく、現場の支援者を支え、子どもを安全に導くために活用される。そして、それは同時にエビデンス based policy making による、データに基づく政治的意思決定の支援につながっていく。本事業における標準 DB の構築は、必要性のある予算根拠の明示や人材育成を進め、現場を定量的に評価し、変革していくための武器となる。データベースというリソースは今後子どもの安全を最大化に寄与する重要な要素の一角を担うものとなるだろう。

国の方向として、実務と研究、そして政策評価に活きる DB 設計を拡張性として組み込み、子どもの安全に直結する DB 利活用をイメージとして持って頂ければ、幸いである。

巻末資料

1. NCANDS データベース例

2種類あるNCANDSのデータベースのうち、②個人情報がマスクされた上で概要を把握できるデータベースの例を表に示す。

以下には、②の児童ファイルコードブック版の領域番号、Name、名称、ラベルについて以下に掲載する(詳細のカテゴリコードは未掲載)。

表: NCANDS データベース例

領域番号	Name	名称	ラベル
1	提出年度	SUBYR	提出年度
2	州・属領	STATERR	州・属領
3	通告ID	RPTID	通告ID
4	児童ID	CHID	児童ID
5	通告を受けた郡(County)	RPTCNTY	報告書の郡
6	通告日(コールセンターなどの通告日)	RPTDT	通告日
7	調査開始日(CPSによる直接目視による現認日)	INVDAT	調査開始日
8	通告元	RPTSRC	通告元
9	通告に対する処遇決定	RPTDISP	通告に対する処遇決定
10	通告に対する処遇決定日	RPTDISDT	通告に対する処遇決定日
11	通知	NOTIFS	通知

12	通告時の児童年齢	CHAGE	通告時の児童年齢
13	児童の生年月日	CHBDATE	児童の生年月日
14	児童の性別	CHSEX	児童の性別
15	児童の人種アメリカンインディアンまたはアラスカ先住民	CHRACAI	児童の人種アメリカンインディアンまたはアラスカ先住民
16	児童の人種アジア人	CHRACAS	児童の人種アジア人
17	児童の人種黒人またはアフリカ系米国人	CHRACBL	児童の人種黒人またはアフリカ系米国人
18	児童の人種ハワイ先住民またはその他の太平洋諸島の住民	CHRACNH	児童の人種ハワイ先住民またはその他の太平洋諸島の住民
19	児童の人種白人	CHRACWH	児童の人種白人
20	人種判定不能	CHRACUD	人種判定不能
21	児童の民族分類	CHETHN	児童の民族分類
22	居住郡	CHCNTY	居住郡
23	住居形態	CHLVNG	住居形態
24	軍に所属している家族	CHMIL	軍に所属している家族
25	被害歴	CHPRIOR	被害歴
26	一つ目の虐待種別(訳者注主たる虐待種別)	CHMAL1	一つ目の虐待種別

27	一つ目の虐待に対する処遇レベル	MAL1LEV	一つ目の虐待に対する処遇レベル
28	二つ目の虐待種別	CHMAL2	二つ目の虐待種別
29	二つ目の虐待に対する処遇レベル	MAL2LEV	二つ目の虐待種別
30	三つ目の虐待種別	CHMAL3	三つ目の虐待種別
31	三つ目の虐待に対する処遇レベル	MAL3LEV	三つ目の虐待に対する処遇レベル
32	四つ目の虐待種別	CHMAL4	四つ目の虐待種別
33	四つ目の虐待に対する処遇レベル	MAL4LEV	四つ目の虐待に対する処遇レベル
34	虐待死	MALDEATH	虐待死
35	アルコール乱用-児童	CDALC	アルコール乱用-児童
36	薬物乱用-児童	CDDRUG	薬物乱用-児童
37	知的障害-児童	CDRTRD	知的障害-児童
38	情緒障害-児童	CDEMOTNL	情緒障害-児童
39	視覚障害または聴覚障害-児童	CDVISUAL	視覚障害または聴覚障害-児童
40	学習障害-児童	CDLEARN	学習障害-児童
41	身体障害-児童	CDPHYS	身体障害-児童
42	問題行動-児童	CDBEHAV	問題行動-児童

43	その他の病状-児童	CDMEDICL	その他の病状-児童
44	アルコール乱用-保護者	FCALC	アルコール乱用-保護者
45	薬物乱用-保護者	FCDRU	薬物乱用-保護者
46	知的障害-保護者	FCRTRD	知的障害-保護者
47	情緒障害-保護者	FCEMOTNL	情緒障害-保護者
48	視覚障害または聴覚障害-保護者	FCVISUAL	視覚障害または聴覚障害-保護者
49	学習障害-保護者	FCLEARN	学習障害-保護者
50	身体障害-保護者	FCPHYS	身体障害-保護者
51	その他の病状-保護者	FCMEDICL	その他の病状-保護者
52	家庭内における暴力	FCVIOL	DV
53	不適切な住居環境	FCHOUSE	不適切な住居環境
54	経済的問題	FCMONEY	経済的問題
55	公的支援	FCPUBLIC	公的支援
56	処遇決定後の支援	POSTSERV	処遇決定後の支援
57	支援開始日	SERVDATE	支援開始日
58	家族サポート支援	FAMSUP	家族サポート支援

59	在宅支援	FAMPRES	在宅支援
60	里親制度	FOSTERCR	里親制度
61	保護日	RMVDATE	保護日
62	少年裁判所への申し立て	JUVPET	少年裁判所への申し立て
63	申し立て日	PETDATE	申し立て日
64	裁判所指定代理人	COCHREP	裁判所指定代理人
65	養子縁組支援	ADOPT	養子縁組支援
66	ケースマネジメントによる支援	CASEMANG	ケースマネジメントによる支援
67	カウンセリングによる支援	COUNSEL	カウンセリングによる支援
68	保育による支援-児童	DAYCARE	保育による支援-児童
69	教育及び研修による支援	EDUCATN	教育及び研修による支援
70	就労支援	EMPLOY	就労支援
71	家族計画支援	FAMPLAN	家族計画支援
72	健康関連及び在宅医療による支援	HEALTH	健康関連及び在宅医療による支援
73	在宅生活支援	HOMEBASE	在宅生活支援
74	住居支援	HOUSING	住居支援

75	自立生活及び移行生活支援	TRANSLIV	自立生活及び移行生活支援
76	情報提供による支援	INFOREF	情報提供による支援
77	法的支援	LEGAL	法的支援
78	メンタルヘルス支援	MENTHLTH	メンタルヘルス支援
79	若い両親のための妊娠／育児支援	PREGPAR	若い両親のための妊娠／育児支援
80	レスパイトケア[短期入所]による支援	RESPITE	レスパイトケアによる支援
81	特別支援-障害者	SSDISABL	特別支援-障害者
82	特別支援-未成年非行者	SSDELINQ	特別支援-未成年非行者
83	薬物乱用に関する支援	SUBABUSE	薬物乱用に関する支援
84	移動のための支援	TRANSPRT	移動のための支援
85	その他の支援	OTHERSV	その他の支援
86	担当者 ID	WRKRID	担当者 ID
87	スーパーバイザーID	SUPRVID	スーパーバイザーID
88	第一の加害者 ID	PERIID	第一の加害者 ID
89	第一の加害者-関係性	PERIREL	第一の加害者-関係性
90	第一の加害者-親の立場	PERIPRNT	第一の加害者-親の立場

91	第一の加害者-養育者としての立場	PERICR	第一の加害者-養育者としての立場
92	第一の加害者-通告時の年齢	PERIAGE	第一の加害者-通告時の年齢
93	第一の加害者-性別	PERISEX	第一の加害者-性別
94	第一の加害者-人種アメリカンインディアン/アラスカ先住民	PIRACAI	第一の加害者-人種アメリカンインディアン/アラスカ先住民
95	第一の加害者-人種アジア人	PIRACAS	第一の加害者-人種アジア人
96	第一の加害者-人種黒人またはアフリカ系米国人	PIRACBL	第一の加害者-人種黒人またはアフリカ系米国人
97	第一の加害者-人種ハワイ先住民またはその他の太平洋諸島の住民	PIRACNH	第一の加害者-人種ハワイ先住民またはその他の太平洋諸島の住民
98	第一の加害者-人種白人	PIRACWH	第一の加害者-人種白人
99	第一の加害者-人種判定不能	PIRACUD	第一の加害者-人種判定不能
100	第一の加害者-民族分類	PERIETHN	第一の加害者-民族分類
101	第一の加害者-軍人	PERIMIL	第一の加害者-軍人
102	第一の加害者-虐待歴のある者	PERIPIOR	第一の加害者-虐待歴のある者
103	第一の加害者虐待-1	PERIMAL1	第一の加害者虐待-1
104	第一の加害者虐待-2	PERIMAL2	第一の加害者虐待-2
105	第一の加害者虐待-3	PERIMAL3	第一の加害者虐待-3

106	第一の加害者虐待-4	PER1MAL4	第一の加害者虐待-4
107	第二の加害者 ID	PER2ID	第二の加害者 ID
108	第二の加害者-関係性	PER2REL	第二の加害者-関係性
109	第二の加害者-親の立場	PER2PRNT	第二の加害者-親の立場
110	第二の加害者-養育者としての立場	PER2CR	第二の加害者-養育者としての立場
111	第二の加害者-通告時の年齢	PER2AGE	第二の加害者-通告時の年齢
112	第二の加害者-性別	PER2SEX	第二の加害者-性別
113	第二の加害者-人種アメリカンインディアン／アラスカ先住民	P2RACAI	第二の加害者-人種アメリカンインディアン／アラスカ先住民
114	第二の加害者-人種アジア人	P2RACAS	第二の加害者-人種アジア人
115	第二の加害者-人種黒人またはアフリカ系米国人	P2RACBL	第二の加害者-人種黒人またはアフリカ系米国人
116	第二の加害者-人種ハワイ先住民またはその他の太平洋諸島の住民	P2RACNH	第二の加害者-人種ハワイ先住民またはその他の太平洋諸島の住民
117	第二の加害者-人種白人	P2RACWH	第二の加害者-人種白人
118	第二の加害者-人種判定不能	P2RACUD	第二の加害者-人種判定不能
119	第二の加害者-民族分類	PER2ETHN	第二の加害者-民族分類
120	第二の加害者-軍人	PER2MIL	第二の加害者-軍人

121	第二の加害者-虐待歴のある者	PER2PIOR	第二の加害者-虐待歴のある者
122	第二の加害者虐待-1	PER2MAL1	第二の加害者虐待-1
123	第二の加害者虐待-2	PER2MAL2	第二の加害者虐待-2
124	第二の加害者虐待-3	PER2MAL3	第二の加害者虐待-3
125	第二の加害者虐待-4	PER2MAL4	第二の加害者虐待-4
126	第三の加害者 ID	PER3ID	第三の加害者 ID
127	第三の加害者-関係性	PER3REL	第三の加害者-関係性
128	第三の加害者-親の立場	PER3PRNT	第三の加害者-親の立場
129	第三の加害者-養育者としての立場	PER3CR	第三の加害者-養育者としての立場
130	第三の加害者-通告時の年齢	PER3AGE	第三の加害者-通告時の年齢
131	第三の加害者-性別	PER3SEX	第三の加害者-性別
132	第三の加害者-人種アメリカンインディアン/アラスカ先住民	P3RACAI	第三の加害者人種アメリカンインディアン/アラスカ先住民
133	第三の加害者-人種アジア人	P3RACAS	第三の加害者-人種アジア人
134	第三の加害者-人種黒人またはアフリカ系米国人	P3RACBL	第三の加害者-人種黒人またはアフリカ系米国人
135	第三の加害者-人種ハワイ先住民またはその他の太平洋諸島の住民	P3RACNH	第三の加害者-人種ハワイ先住民またはその他の太平洋諸島の住民

136	第三の加害者-人種白人	P3RACWH	第三の加害者-人種白人
137	第三の加害者-人種判定不能	P3RACUD	第三の加害者-人種判定不能
138	第三の加害者-民族分類	PER3ETHN	第三の加害者-民族分類
139	第三の加害者-軍人	PER3MIL	第三の加害者-軍人
140	第三の加害者-虐待歴のある者	PER3PIOR	第三の加害者-虐待歴のある者
141	第三の加害者虐待-1	PER3MAL1	第三の加害者虐待-1
142	第三の加害者虐待-2	PER3MAL2	第三の加害者虐待-2
143	第三の加害者虐待-3	PER3MAL3	第三の加害者虐待-3
144	第三の加害者虐待-4	PER3MAL4	第三の加害者虐待-4
145	AFCARSID(The Adoption and Foster Care Analysis and Reporting System)	AFCARSID	AFCARSID
146	虐待発生日	INCIDDT	虐待発生日
147	通告時刻	RPTTM	通告時刻
148	調査開始時刻	INVSTRM	調査開始時刻
149	死亡日	DEATHDT	死亡日
150	社会的養護終結日	FCDCHDT	社会的養護終結日

2. 三重県データを用いた再相談予測モデルの性能(ROC 曲線)

機械学習の予測性能を示す指標には、複数の指標が存在する。

感度とは、「実際に再相談である事例を、何パーセント“再相談である”と予測できたか」を示す指標であり、値が100%に近いほど「見逃しが少ない」ことを示す。一方、特異度とは「実際に再相談ではない事例を何パーセント“再相談ではない”と予測できたか」を示す指標であり、値が100%に近いほど「再発リスクが低いものを的確に除外できているか」を示す。これら二つの指標はトレードオフの関係にある。具体的には、「全て再相談であると予測」した場合は高い感度(“再発事例を見逃さない”)、が得られるが、低リスク事例を正しく除外できないために特異度は著しく低下する。

感度と特異度のような、それぞれ目的の異なる指標を総合する評価方法の一つに、ROC曲線の描画と曲線下面積(Area Under the Curve: AUC)がある。ROC曲線とは、縦軸に感度、横軸に1-特異度を取り、閾値を連続的に変化させた時の両者のトレードオフ関係を描画した図のことを指す。AUCとは、ROC曲線下の面積を示し、AUCが大きいほど予測の性能は高いと評価される。

三重県データを用いた再相談予測の結果(20回の解析結果を統合)、曲線下面積は0.842となった。またその95%信頼区間は、0.826から0.857となった。

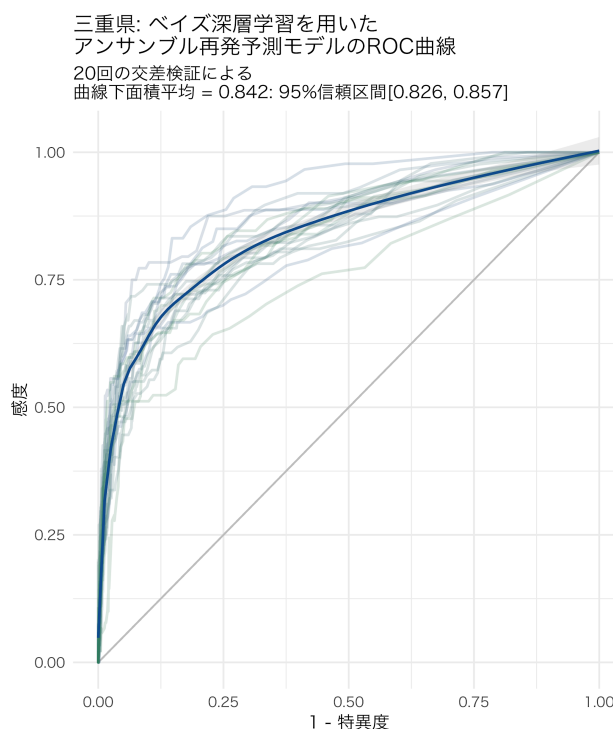


図 三重県: 再相談予測の機械学習モデル性能(ROC 曲線)

3. 神奈川県データを用いた再相談予測モデルの性能

神奈川県データを用いた再相談予測モデルの結果について詳述する。三重県の場合と同様であるが、機械学習の結果では、個別事例に対する「再相談スコア」が算出される。この数値は、0から1までの値を取り、1に近いほど「再相談になるだろう」と予測されていることを示す。このようなスコアを横軸にとり、各0から1までのスコアに該当した事例数を縦軸とするヒストグラムを描画した(下図 左パネル)。この時、各事例が実際に再相談であったか否かを色分けした。非再相談の事例がスコア0の近くに偏っている傾向が見て取れる。

機械学習では、スコア 0.5 以下を「非再相談」とし、スコア 0.5 より大きい場合を「再相談」と判断して予測の結果を利用することが多い。

再相談/非再相談の分類の閾値を 0.5 とした場合の、分類結果が下図右パネルである。横軸を実際の再相談の有無とし、縦軸が予測による再相談の有無となっている。結果は、「実際の再相談を正しく“再相談である”と予測できた件数が検証データで 174 件となっており(左下)、「実際の非再相談を正しく“非再相談である”と予測できた件数が 1689 件となっている。一方で、予測と実際にズレがあった事例も、238 件、161 件と存在していることがわかる。

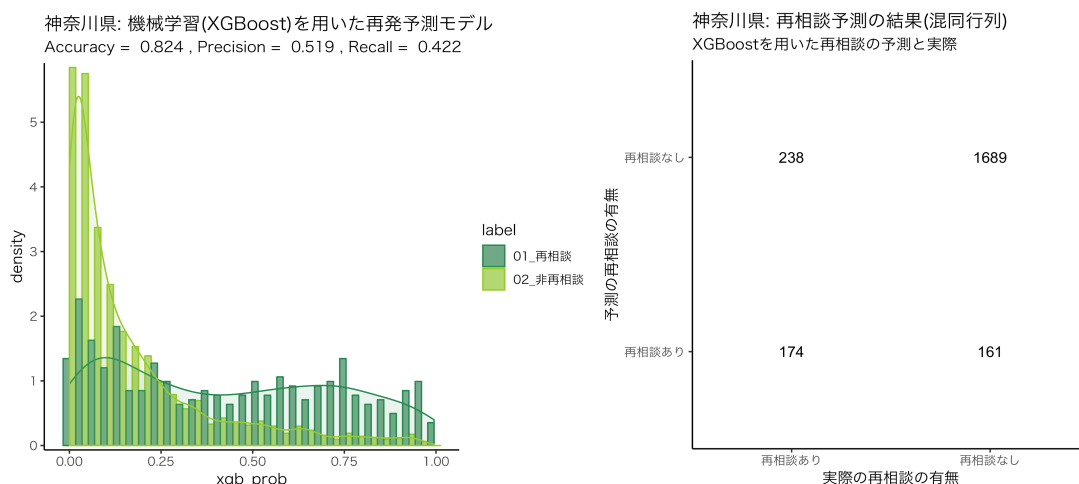


図 神奈川県: 再相談予測モデルの性能一例

感度と特異度の観点から ROC 曲線を描画したところ、その曲線下面積は 0.77 となった。

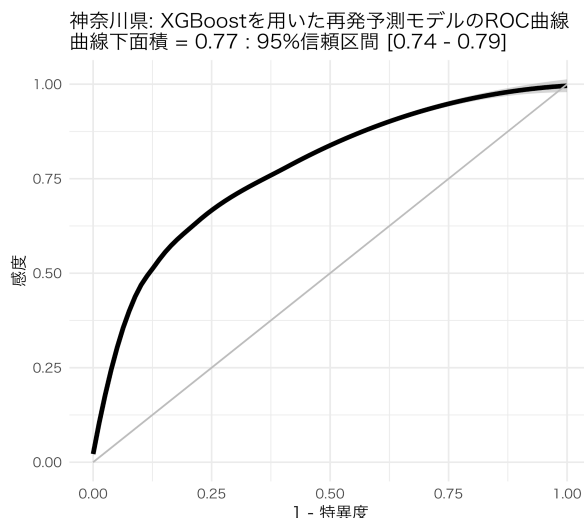


図 神奈川県: 再相談予測の機械学習モデル性能(ROC 曲線)

4. 千葉県データを用いた再相談予測モデルの性能

上記、神奈川県の結果と同様の評価指標を用いた結果が下図である。千葉県のデータでは、スコアが 0 に近い「再相談ではないだろう」という予測と、スコアが 1 に近い「再相談だろう」という予測結果に対して、実際の再相談と非再相談が正しく分類される傾向にある。

ることが見て取れる(下図 左パネル)。すなわち、実際の非再相談事例が横軸の0に近い方向に偏り、反対に実際の再相談が1に近い方向に偏っていることがわかる。ただし、再相談事例に関しては、0に近い値で「再相談ではないだろう」と予測されたが、実際には再相談だったという事例も多い。再相談と非再相談を切り分ける情報が不足したことで、誤って「再相談ではない」と予測してしまっている結果を示している。

予測スコアに対する閾値を0.5として再相談/非再相談の分類を行った結果(下図 右パネル)、「実際の再相談を正しく“再相談である”」と予測できた件数が検証データで656件となっており(左下)、「実際の非再相談を正しく“非再相談である”」と予測できた件数が1652件となっている。一方で、予測と実際にズレがあった事例も、466件、321件と存在していることがわかる。

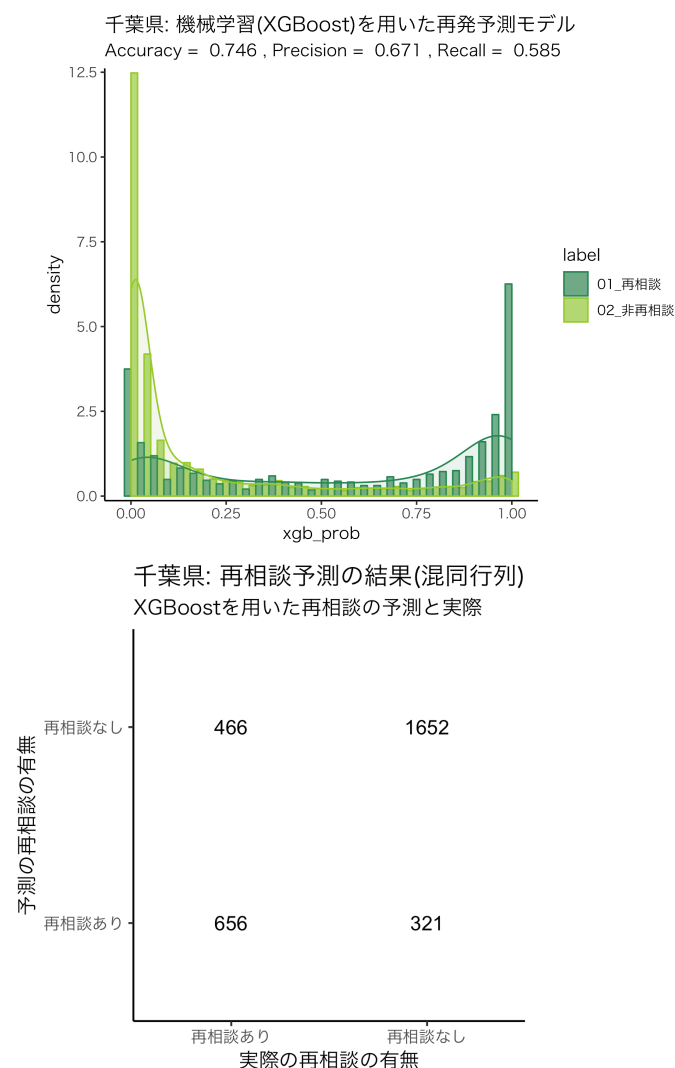


図 神奈川県: 再相談予測モデルの性能一例

また、感度と特異度の観点から ROC 曲線を描画したところ、その曲線下面積は0.78 となった。

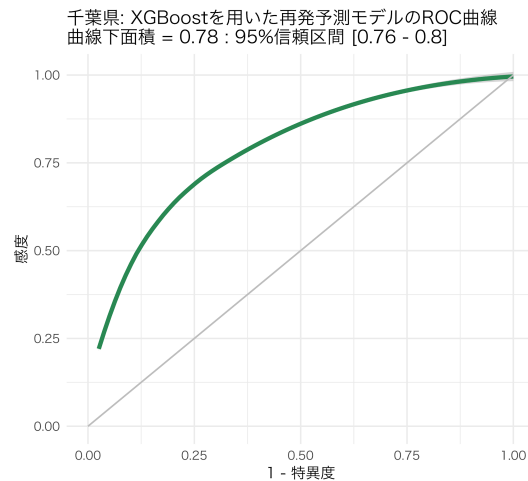


図 千葉県: 再相談予測の機械学習モデル性能(ROC 曲線)

5. 三重県データを用いた対応日数予測モデルの性能

対応日数の予測は、個別事例における「実際の日数」と「予測される日数」の差がどの程度小さいかを総合して評価することが多い。一般には、Mean Square Error や Root Mean Square Error といった指標が用いられることが多い。ここでは、解釈可能性を高めるため実際の日数と予測の日数をカテゴリ化し、その精度(Accuracy)で評価した。

下図に、予測日数区分を横軸とし、実際の日数区分を縦軸とするクロス表(混同行列)を示す。予測と実際の対応関係があり、対角線上に一致して整列していることがわかる。精度は、0.934 となっていた。

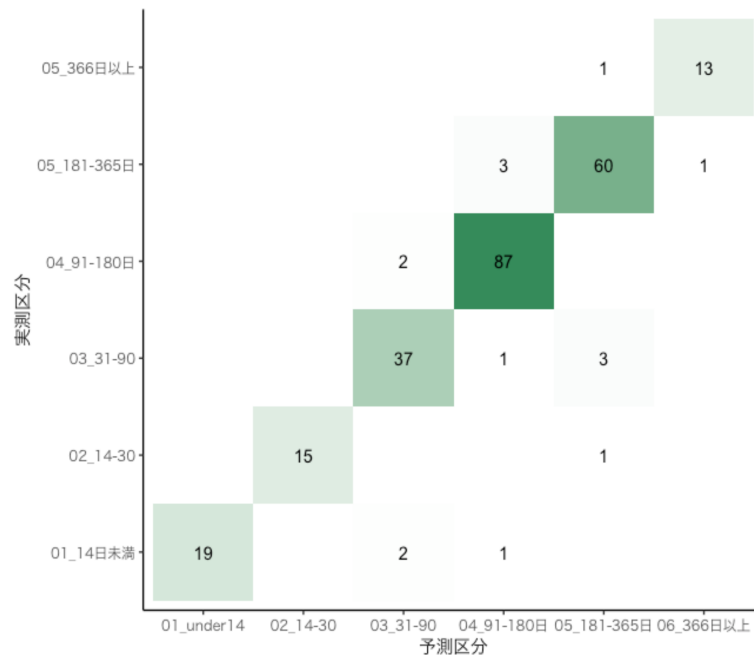


図 三重県: 対応日数予測の機械学習モデル性能

6. 神奈川県データを用いた対応日数予測モデルの性能

三重県と同様にして、神奈川県での解析結果を下図に示す。予測と実際の対応関係がおおよそ示されており、ズレはあるものの対角線上への整列傾向にあることがわかる。予測精度は、0.375となっていた。

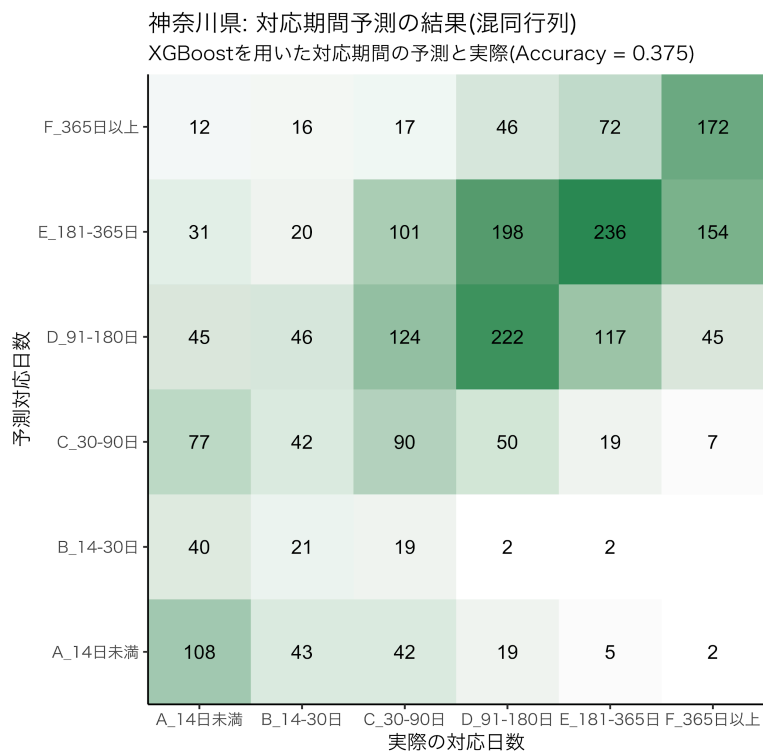


図 神奈川県: 対応日数予測の機械学習モデル性能

7. 千葉県データを用いた対応日数予測モデルの性能

千葉県データによる解析結果を下図に示す。予測と実際の対応関係がおおよそ示されており、ズレはあるものの対角線上への整列傾向にあることがわかる。予測精度は、0.422となっていた。

千葉県: 対応期間予測の結果(混同行列)
 XGBoostを用いた対応期間の予測と実際(Accuracy = 0.422)

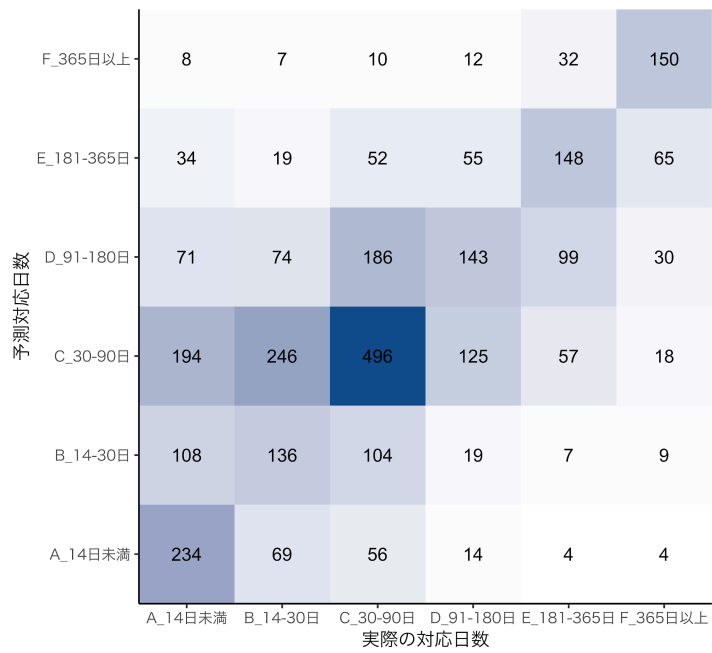
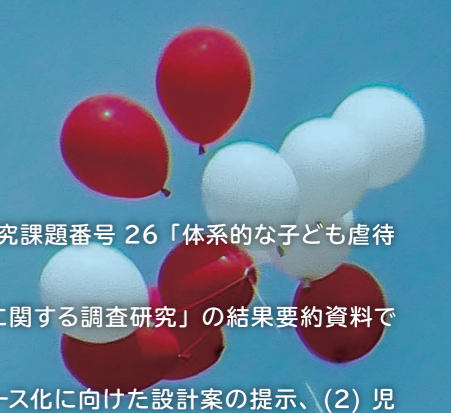


図 千葉県: 対応日数予測の機械学習モデル性能



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。

CONTENTS 目次

- はじめに（背景と目的） P2
- 総合要約 P2
- 第 1 部 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
- 第 2 部 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

総合要約 IMPORTANT DISCOVERY

本事業の主たる目的は次の二点である。

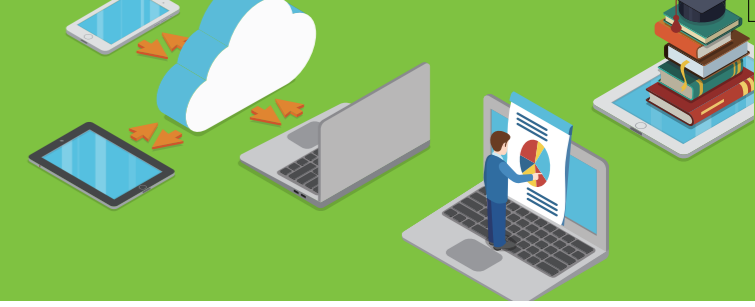
1. 情報集約システムを活用した情報共有や関係機関の連携についての好事例を把握し、当該事例で用いられている情報集約システムが児童虐待の端緒となる要因の把握と、児童虐待の発生予防・早期発見に資すること示すこと。
2. 虐待統計の国際比較等を調査した平成 29 年度報告等を踏まえ、国際比較が可能な調査項目を用いた虐待統計のモデル実施を行うこと。

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、国際比較に耐えう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。



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。

データベース (DB) 整備に向けたプロセス



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、各項目は紙に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。

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、現場に应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ことが利点である。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
- 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
- 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
- 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、データ入力の数カントリールール（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、ペ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。

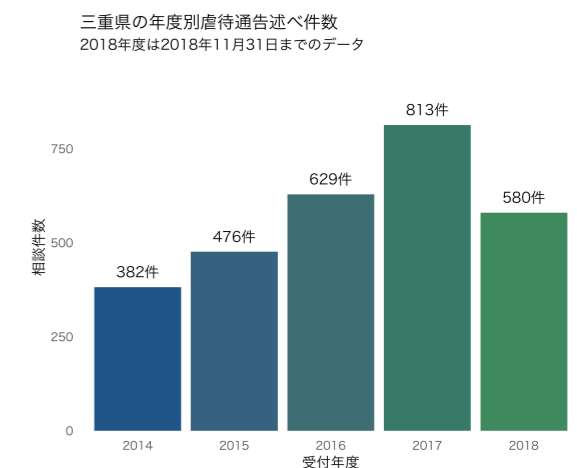
結果

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

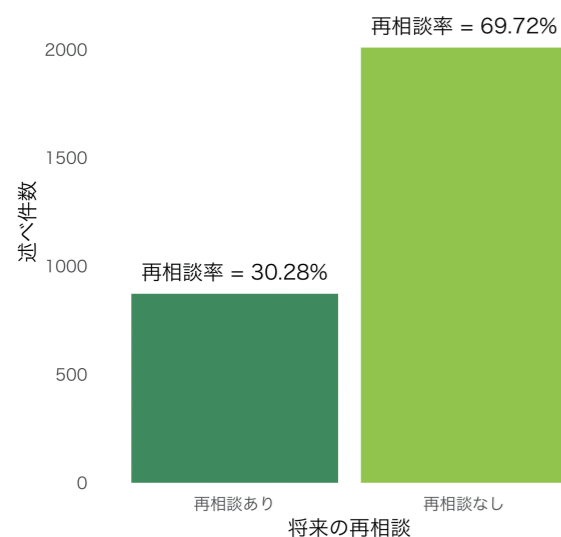
三重県データベースから抽出したデータセットを解析用に結合した結果、解析対象となったのは述べ7395件の虐待相談レコードであった。年度別の虐待通告件数は、2017年度で2022件となっている。2018年度については、11月31日時点までのデータであるため、参考値となる（右図）。

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）。



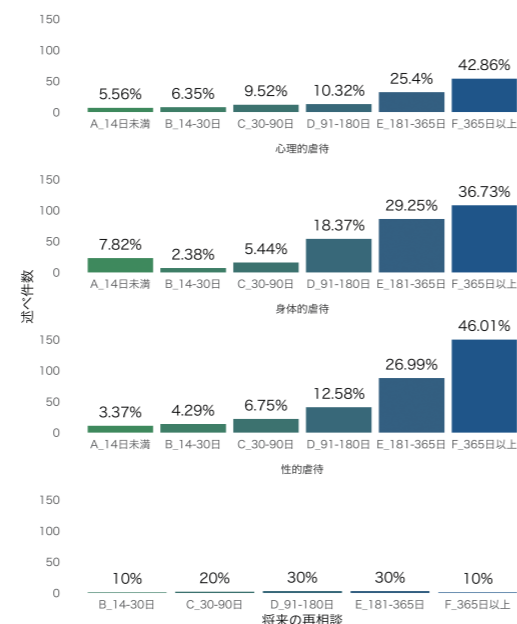
三重県における再相談発生率

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)

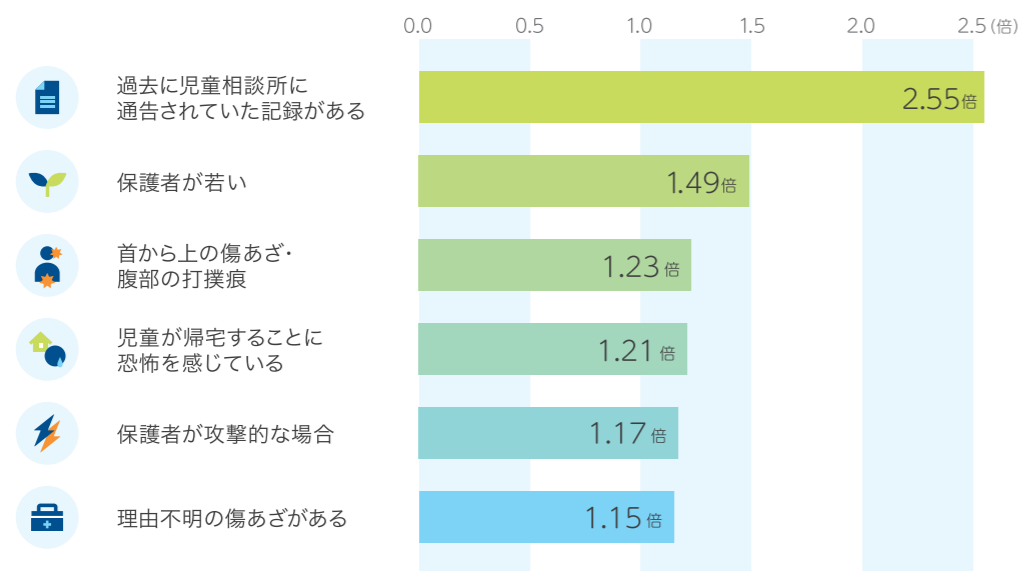


図 三重県：将来の再相談に関わる関連要因 (調整オッズ比)

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



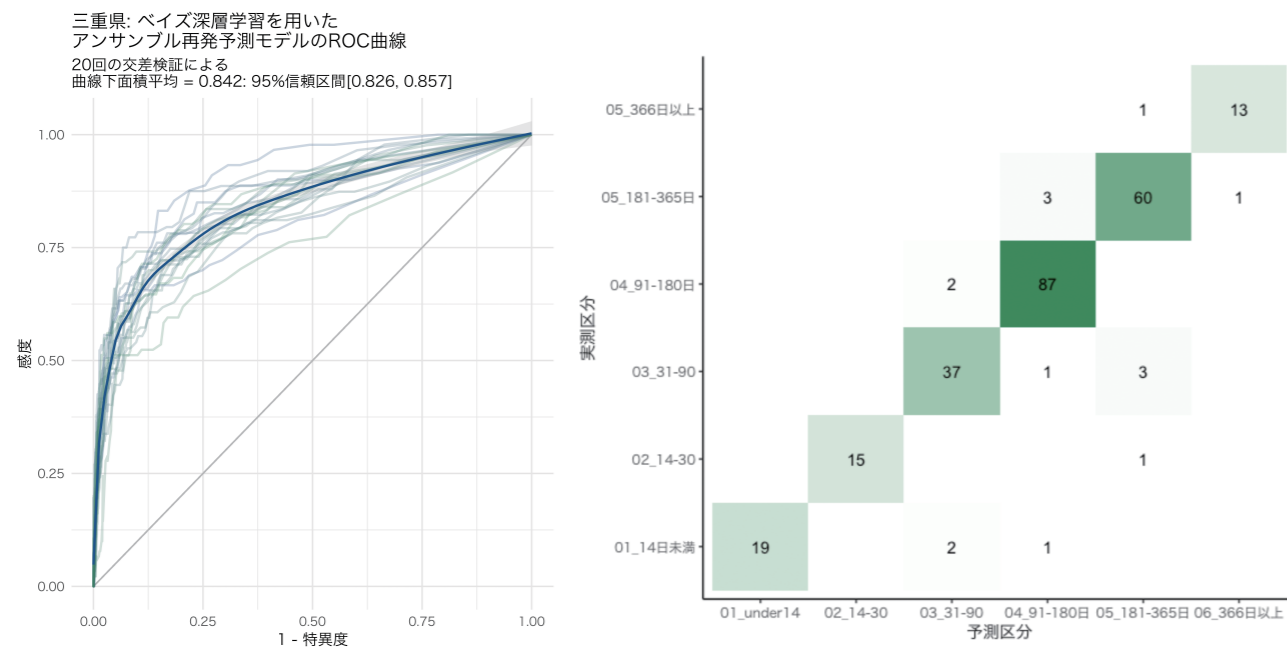
神奈川県・千葉県における長期化関連要因について

神奈川県の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略が定立可能となる。

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が期待される。