

報告 2

日医健診標準フォーマットへの 健診データの変換作業について

日本医師会総合政策研究機構(日医総研)

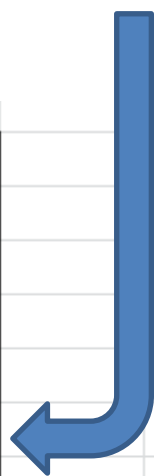
客員研究員 窪寺 健

医師会共同利用施設の健診データ変換状況

変換作業の状況

- 2015年9月15日まで 全パイロット施設に対する協力依頼
2015年10月7日から 全パイロット施設に対しサンプルデータ提供正式依頼
2015年10月21日から データ受領・変換作業開始、一部分析開始
2015年12月22日現在 データ変換ツール作成(赤文字は確認まで終了)
2015年1月より 標準変換ツール配布

	日医共同利用施設対象施設	サンプル件数
1	A 医師会	2,321
2	B 医師会	245
3	C 医師会	1,001
4	D 医師会	1,555
5	E 医師会	4,778
6	F 医師会	1,372
7	G 医師会	1,044
8	H 医師会	274
9	J 医師会	13,197
10	K 医師会	62
パイロット施設15施設中、5施設については繁忙期のため提供が遅れている		25,849



データ変換における具体的な事象

1. 複数の健診機関のデータ変換をしてみても

- ・ 画像検査の「所見区分」と「所見」について設定ルールが統一されていない。
- ・ 食後時間は「* 時間以上」「* * * 分以上」「コード」「項目なし」などが混在する。
- ・ 特定健診以外は「判定」における分類定義(機能・臓器)が統一されていない。
- ・ 数値項目に「テキスト表記」が混在する項目の処理が必要。
→ 視力については「0.10未満(以下)」については「0.00」と表記するルールとした。

しかし、視力以外の項目でルールの設定ルールの検討が必要。

(検査項目)

肝炎検査定量、ヘリコバクターIgG抗体定量、CRP、聴力デシベル、腫瘍マーカー 等

(出現パターン)

「* * 未満」の標記 「<* *」、「↓* *」、「* * ↓」

「* * 以上」、「* * 超」の標記 「* * <」、「* * ↑」、「↑* *」

上記に加えて桁数の不統一

肝炎検査定量は「0.01未満」や「1.0未満」など

以上により、対象とする健診機関を増やすとともにこれらの事象は拡大していくことが予想される。いずれかのタイミングで表記方法などの標準化が必要となる。

データ変換における具体的な事象

2. 二次データベースからデータを抽出すると

(保険者システムあるいは複数健診機関から収集したデータベースからの抽出)

- ・ 血球計算の単位($/\mu\text{l}$ 、 $10^3/\mu\text{l}$ 、 $10^2/\mu\text{l}$ 、 $/10^4$ 、 $/10^6$)が混在している。
- ・ 画像検査等に左右を区別する項目がないので、左右情報が取れない。
- ・ 臨床検査項目等が複数の検査方法が含まれている可能性が疑われる。

(原因と思われること)

- ・ 本来、保険者システムは請求処理を目的として構築されており、蓄積された健診データの分析を目的とはしていない特徴がある。
- ・ データ仕様が異なる複数の健診機関から健診データを収集している場合には十分な対応がされていなければ、途中に介在する変換プログラムの仕様がブラックボックスとなり、単位がバラバラのものを同一項目内に格納している可能性もある。

以上により、複数の健診機関から健診データを集約しているシステムからデータ収集を行う場合には健診機関個々の由来情報の収集はできない。データをクリーニングできないデータベースの二次利用には疑問がある。

データベース構築に向けて

1. 異なる健診機関のデータを集約し蓄積を行うには

①【流通・相互運用性】

健診機関の運用情報を伴う由来情報を流通をすべき。

健診機関やベンダーは由来情報についての重要性を認識すべき。

②【流通・真正性】

健診システムが保持するデータはそのままの形で流通されるべき。

元データを変換するなら統一された変換ルールで行われるべき。

③【施設内運用】

健診機関の所見や判定は独自ルールではなく、標準化されていくべき。

健診機関における検査等運用を健診コース別に保持すべき。

④【業界統一】

健診システムにおけるデータ管理項目は標準化されるべき。

検査項目以外の顧客要望項目の設定ルールは標準化されるべき。

2. 日医健診標準フォーマット対応 **変換ツールが普及する条件**

- ① 健診システムの全管理項目をそのままの形で出力したものを、簡単な操作で日医健診標準フォーマットへの変換が可能であること。
- ② 画像検査における「所見区分」はじめ、特定健診における請求根拠理由項目、労基署報告で利用する報告項目等のユースケースに対応できること。
- ③ 関係省庁と綿密な連携を行い、健診団体との意思疎通が行われること。
- ④ 健診団体総意のもと健診データ管理項目の統一を決定し、システム開発ベンダーに共有すること。
- ⑤ 医療の質の向上に貢献できるデータベース構築を行い、科学的根拠を創出すること。
- ⑥ 受診者の同意のもと、健診データは「かかりつけ医」が利用することとし、公的な医療関係団体が二次利用できる環境整備を行うこと。
- ⑦ 二次利用において新たなビジネスモデルを作り、健診団体としてデータベース運用を長期的に可能にすること。

医師会共同利用施設の健診データ変換のステップ

Step 1 (施設内DB構築まで)

各施設で日医健診標準フォーマットによるデータファイルの作成。



①健診システムで
CSVファイル作成。

②外付けHDにある「
標準化ツール」を起動
。

③「標準化ツール」で
CSVファイルを標準フォ
ーマットに変換。

④変換後のデータファイルは
外付けHDに蓄積。

Step2

⑤クラウド化するには
外付けHDから取り出し、
国際標準・国内標準に対応し
た共用DBの構築を行う。

変換ツール②によるJAHIS、ISO13606への対応

国立研究開発法人日本医療機構(AMED) 平成27年度 臨床研究等 ICT 基盤構築研究事業 採択案件
 「全国共同利用型国際標準化健康・医療情報の収集及び利活用に関する研究」「千年カルテプロジェクト」
 (連携先: テクノロジックアート、特定非営利活動法人 日本医療ネットワーク協会)

分類	学校検診項目	コンセプト名			
①身長、体重及び座高	身長	Height/Length			
	体重	Body weight			
	座高	Measurement of Body Segment			
②栄養状態	診				
③脊柱及び胸部の疾病及び異常の有無	他	区分名称	JAHIS(XML表示名)	単位	項目コード(17桁)
		身体計測	身長	cm	9N001000000000001
		身体計測	体重	kg	9N006000000000001
④視力及び聴力	視	医師の判断	医師の診断(判定)	文字列	9N511000000000049
		医師の判断	医師の意見	文字列	9N521000000000049
		医師の判断	医師の診断(判定)	文字列	9N511000000000049
		医師の判断	医師の意見	文字列	9N521000000000049
		がん検診・生体検査等	視力(左)	数字	9E160162200000001
		がん検診・生体検査等	視力(左矯正)	数字	9E160162600000001
	聴	がん検診・生体検査等	視力(右)	数字	9E160162100000001
		がん検診・生体検査等	視力(右矯正)	数字	9E160162500000001
		がん検診・生体検査等	聴力(左:1000Hz)	コード	9D100163500000011
		がん検診・生体検査等	聴力(左:4000Hz)	コード	9D100163600000011
		がん検診・生体検査等	聴力(右:1000Hz)	コード	9D100163100000011
		がん検診・生体検査等	聴力(右:4000Hz)	コード	9D100163200000011
		がん検診・生体検査等	聴力(検査方法)	コード	9D100164000000011
⑤眼の疾病及び異常の有無	結	がん検診・生体検査等	聴力(その他の所見)	文字列	9D100160900000049
		医師の判断	医師の診断(判定)	文字列	9N511000000000049
⑥耳鼻咽喉頭疾患及び皮膚疾患の有無	他	医師の判断	医師の意見	文字列	9N521000000000049
		歯			
⑦歯及び口腔の疾病及び異常の有無	診	医師の判断	医師の診断(判定)	文字列	9N511000000000049
		医師の判断	医師の意見	文字列	9N521000000000049
⑧結核の有無	診	医師の判断	医師の診断(判定)	文字列	9N511000000000049
		医師の判断	医師の診断(判定)	文字列	9N511000000000049
⑨心臓の疾病及び異常の有無	心	医師の判断	歯科医師による健康診断	文字列	9N531000000000049
		医師の判断	歯科医師の意見	文字列	9N541000000000049
		医師の判断	医師の診断(判定)	文字列	9N511000000000049
		医師の判断	医師の診断(判定)	文字列	9N511000000000049
		がん検診・生体検査等	心電図(所見の有無)	コード	9A110160700000011

変換後データの一次分析結果(抜粋・暫定集計値)

1. 由来情報の集計結果

- ①<体重> 測定値の調整条件が標準化されていない
- ②<HbA1c> 施設内基準範囲と「共用基準範囲」

2. 変換後データの分布

- ①<腹囲> 特定健診判定基準と男女実測値の特徴
- ②<血圧> 施設における測定単位の特徴
- ③<アルブミン> 施設ごとのピークや高値のバラつき
- ④<クレアチニン> 結果データの単位の問題

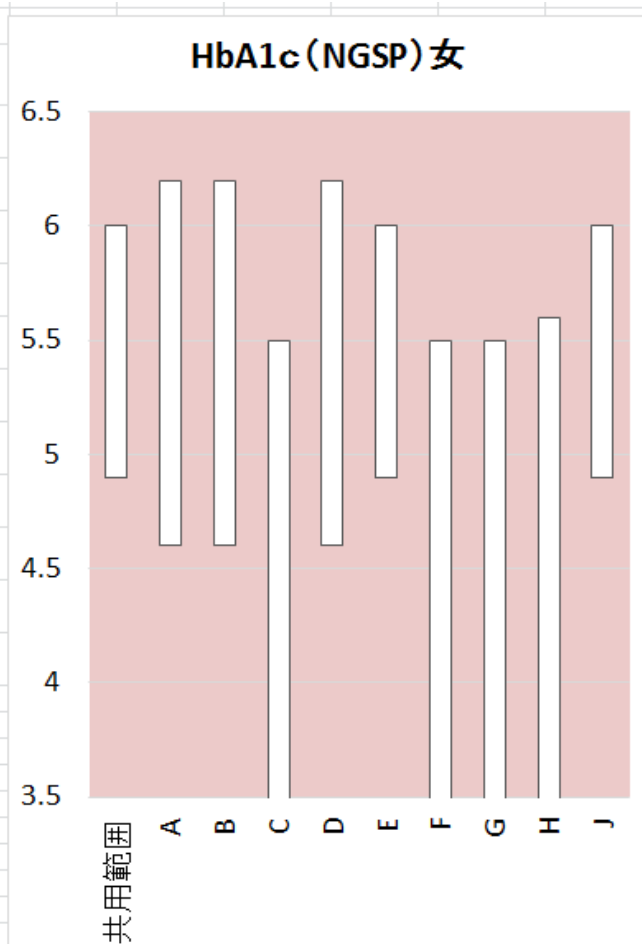
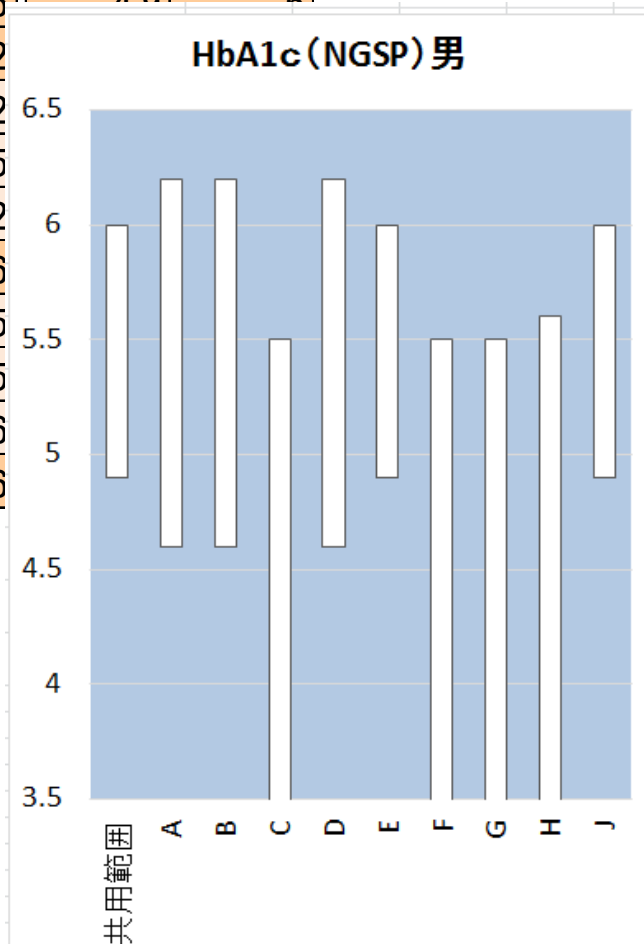
由来情報の集計結果：体重測定機器の調整方法

施設	体重測定機器	体重の調整方法(実測値との調整)
A	TANITA DC-250	一律 -1kg
B	TANITA WB-150	実施医療機関による
C	TANITA DC-250	?
D	TANITA TBF-210	健診着着用時 -500g
E	U-well2	健診着着用
F	TANITA DC-250	健診着着用時 -500g、平服時 -800g
G	KS-197Dp (関西精機)	夏季 -500g、冬季 -1kg
H	?	一律 -500g
J	TANITA TBF-210	?

体重測定においては着衣による実測値からの調整方法にバラつきがあり、調整する方法も様々である

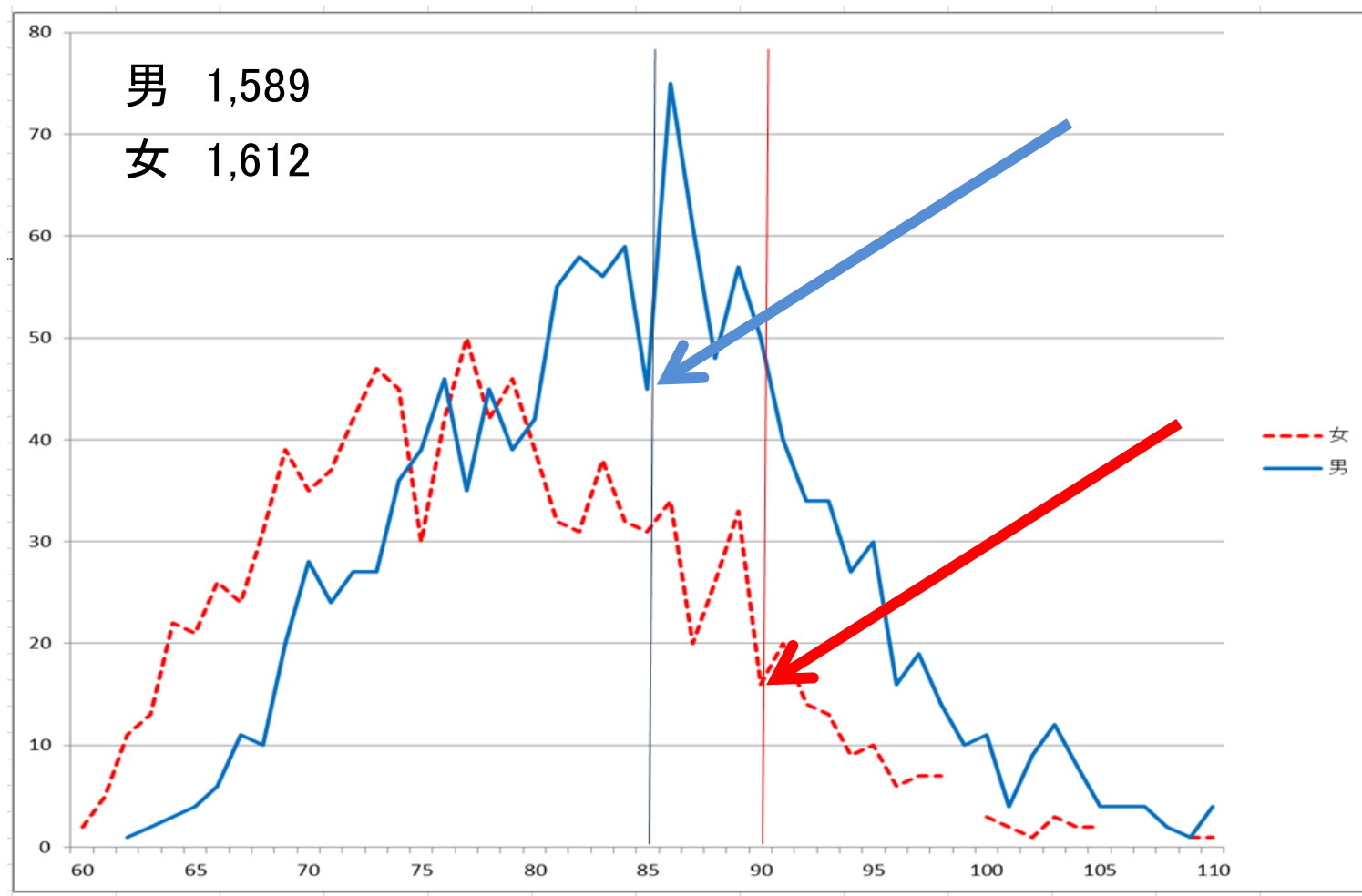
由来情報の集計結果:HbA1cの基準範囲

HbA1c	男/下限	男/上限	女/下限	女/上限
共用範囲	4.9	6.0	4.9	6.0
A	4.6	6.2	4.6	6.2
B	4.6	6.2	4.6	6.2
C	5.0	5.5	5.0	5.5
D	4.6	6.2	4.6	6.2
E	4.9	6.0	4.9	6.0
F	5.0	5.5	5.0	5.5
G	5.0	5.5	5.0	5.5
H	5.0	5.6	5.0	5.6
J	4.9	6.0	4.9	6.0



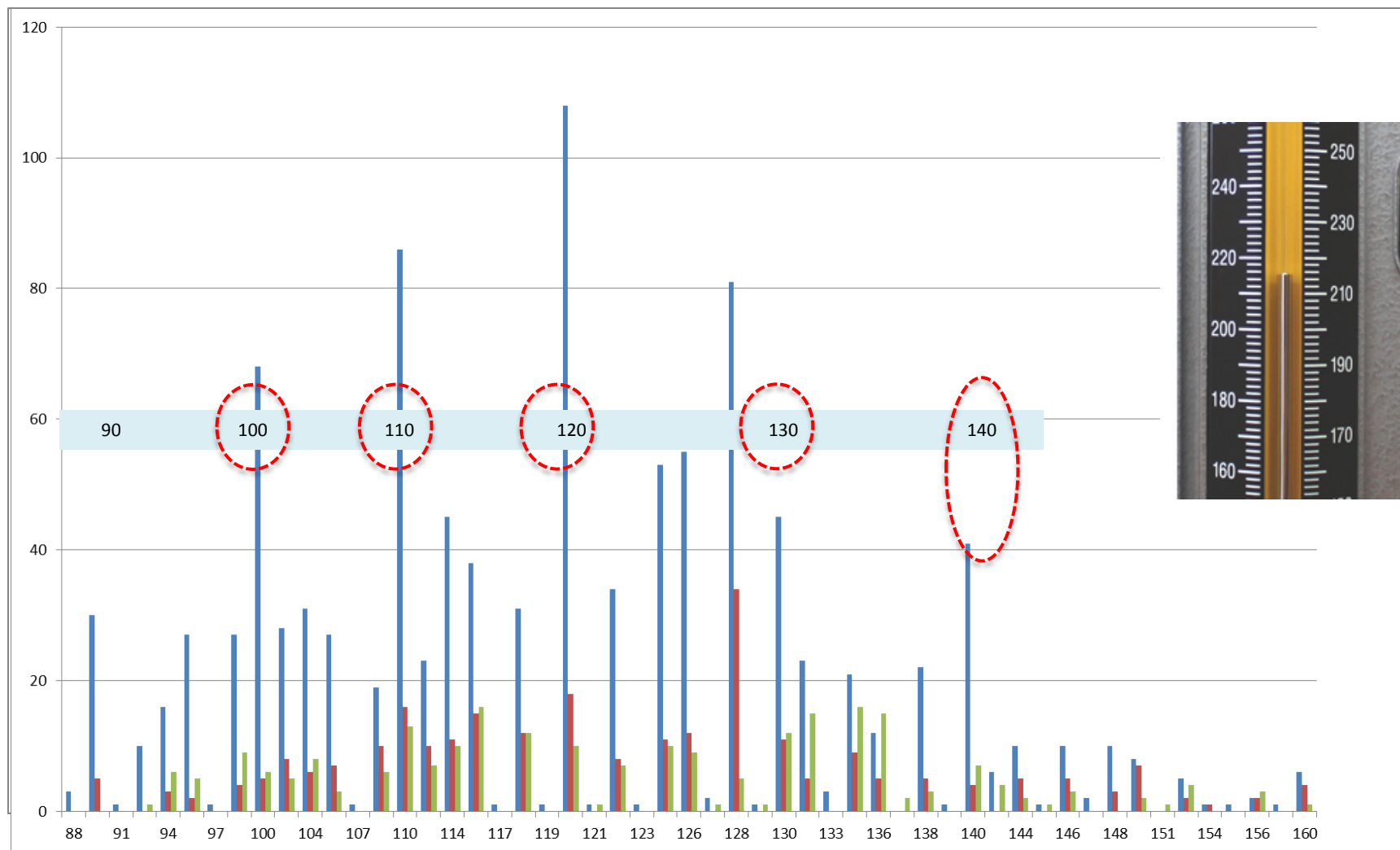
日本臨床検査標準化協議会(JCCLS)が示しているHbA1cの共用基準範囲は4.9~6.0%(NGSP)であるが、各施設の基準範囲はバラついている

収集データの分布: 腹囲



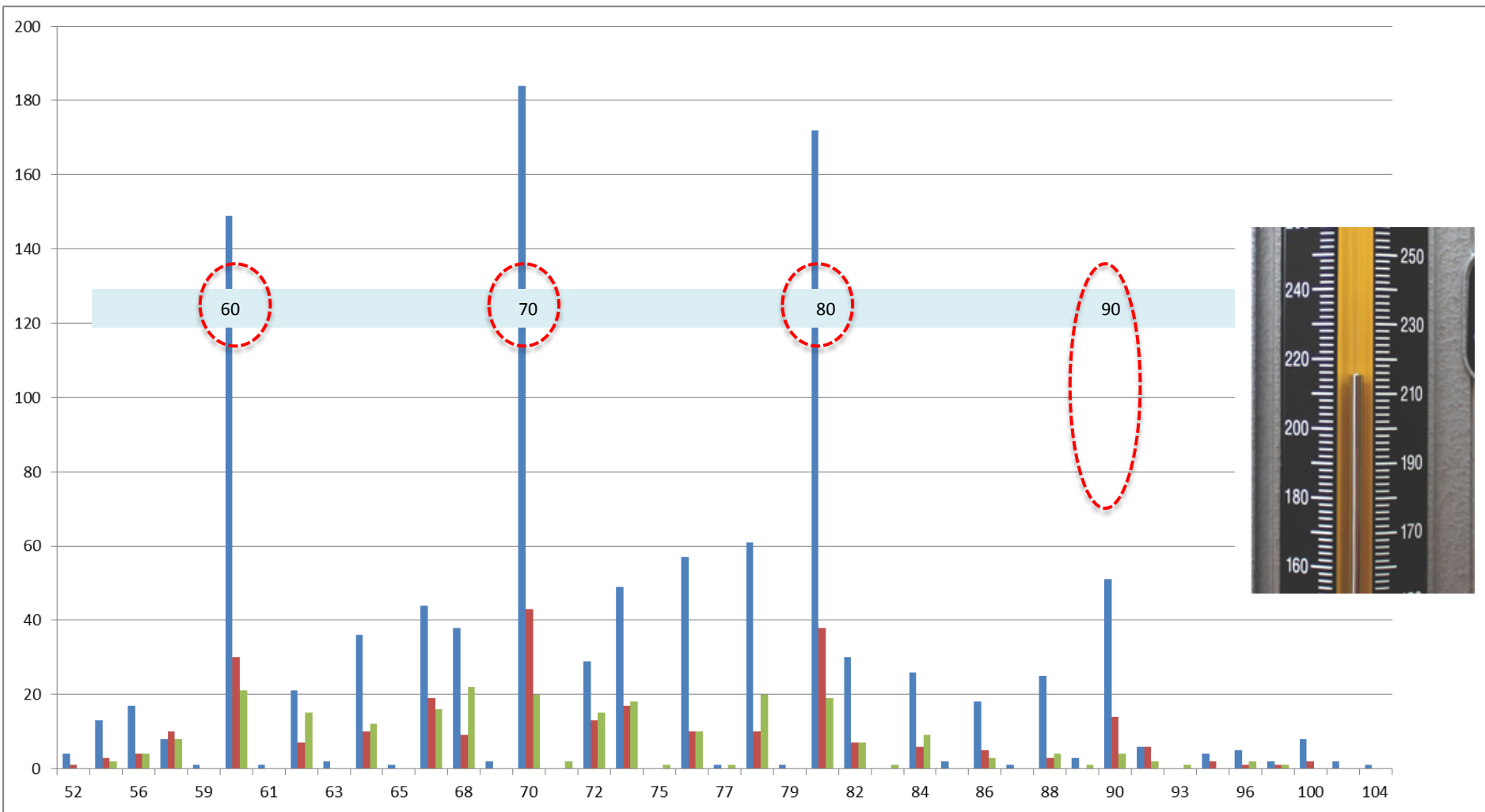
男性女性ともに、特定健診階層化判定リスクとなる85cm、90cmが落ち込んでおり、男性は「リスク要因側」に、女性は「リスク要因回避側」に偏る傾向がみられる。

収集データの分布: 収縮期血圧



1施設では100mmHg、110mmHg、120mmHgと10単位ごとの結果が突出している

収集データの分布: 拡張期血圧



1施設では60mmHg、70mmHg、80mmHgと10単位ごとの結果が突出して多い

健診データの共用DB構築作業(再掲)

1. 共用DB構築の目的

- ①全国の健診機関に対する日医健診標準フォーマットへの変換ツールの配布にあたり、本ツールを利用し、異なる健診システムや各種健診・検診の健診データが「日医健診標準フォーマット」という形式の「ひとつのデータベース」に格納できることを明らかにすることで、広く利用を周知する。
- ②共用DBに格納された健診データを集計・分析することで、特定健診、一般定期健診、がん検診、人間ドック等で実施されている検査項目の把握や、項目ごとの由来情報の特性を明らかにして今後の健診データの標準化の推進に役立てる。また、所見類の標準化案の検討を行う。

2. 共用DBからの集計データ収集の対象施設

①医師会共同利用施設

変換ツール配布の準備として、サンプルデータを提供いただいた15の医師会共同利用施設のうち、過去データの一括抽出と匿名化処理によるデータの変換作業が可能な施設

②全国規模の健診機関

広域で健診を実施している医療法人等の健診機関で「前項」の作業が可能な施設(4法人からデータ提供)

健診データの共用DB構築作業状況

日医健診標準フォーマットへのデータ変換状況：健診区分単位

	施設システム	特定健診	事業者健診	がん検診	人間ドック	
	抽出CSVデータ数	協会けんぽ	一般定期健	対策型		
施設A	2,321		2,321			
施設B	245				245	
施設C	1,001	200	801			
施設D	1,555		1,177	178	200	
施設E	4,778		4,120	81	577	
施設F	1,372	222	910		240	
施設G	1,044	68	976			
施設H	274	132	82		60	
施設J	13,197	7,474	4,278	706	739	
施設K	62					詳細未
施設L	104,718					詳細未
施設M	648,145					詳細未
施設N	966				966	
施設O	761				761	
施設P	43,048			20,012	23,036	
合計	823,487	8,096	14,665	20,977	26,824	

各健診区分の健診データが共用DBに格納されることが確認された