Using the Checkmate Macro

Checkmate is a macro in an Excel spreadsheet developed by the REDCap team at the University of Colorado – Anschutz Medical Campus. It scans data dictionaries for keywords, such as birth or address, and marks them as potential identifiers. It also scans for text fields that may benefit from additional validation (e.g., integer/number, min/max). Running the macro generates a report in an Excel spreadsheet that can be used to verify that identifiers are marked and other best practices are followed.

Checkmate uses a conservative approach to marking identifiers. This means any field containing a keyword may be marked as an identifier, such as questions containing keywords like "name" (e.g., "What is the name of the organization?"). Researchers are to use their best judgement when interpreting Checkmate results and applying it to their project. The IHSI REDCap team will continue to perform their own check of identifiers before moving any project into production.

Checkmate *does not* automatically update data dictionaries. Any necessary changes must be made by either updating the data dictionary and uploading it into the REDCap project or editing the project through the Online Designer within REDCap.

Email the IHSI REDCap team for a copy of the Checkmate macro Excel file.

Instructions

1. Download your most current data dictionary and open the file (see How to Use a Data Dictionary for more information).
2. Open the "Checkmate" Excel file.
3. Copy your entire data dictionary and paste it into the "Checkmate" Excel file at row 20.

16	8. One of more of the values in column F [Choices, Calculations, OR Slider Labels] is not consistant through out the data dictiona									
17											
18	Fields in green meet best practice guidelines										
19											
20	Variable / Field Name	Form Nam	Section He	Field Type	Field Label	Choices, C	Field Note	Text Valid	Text Valid	Text Valid	Identifie
21	study_id	consent_form		text	Study ID						
22	research_study	consent_form		radio	Welcome	1, I consent | 0, I do not consent					

4. While in the "Checkmate" Excel sheet, press "Cntrl + q" to run the macro.
   a. Please note that if you are using a university computer, you may need to contact your local IT support to enable the ability to run macros on your computer.
5. Once the macro is complete, results will be displayed in the blue "General Recommendations" rows below your data dictionary and in Columns S and T.
   a. *General Recommendations* – These rows will show the percent and number of fields that are in free form text and the number of forms that have more than 30 fields.
   b. *Column S* – This is the "Best Practice (BP) Suggestions" column. The numbers listed within this column correspond to the "Specific Recommendations" listed in the yellow rows above and below your data dictionary. Additionally, the corresponding data dictionary column (e.g., Text Validation Type OR Show Slider Number) it recommends adjusting for each marked field will be highlighted yellow.

I ILLINOIS
IHSI | Interdisciplinary Health Sciences Institute

      c. ***Column T*** – This is the "Best Practice (BP) Follow" column. Fields marked green meet best practice guidelines. The numbers listed also correspond to the "Specific Recommendations" listed in the yellow rows.

6. Check for possible identifiers ("1" in columns S or T) and confirm if they have been marked as "Identifiers" as needed. Any field that contains a HIPAA identifier *must* be marked as an identifier, even if the research team does not plan to remove that data from REDCap.

7. Make any needed corrections or adjustments to your project. You can do this in two ways:
    a. Update your data dictionary CSV file and upload the new data dictionary to your REDCap project. Please note, you *cannot* upload the Checkmate Macro Excel file into REDCap.
    b. Alternatively, you can make changes in REDCap by using the "Edit Field" function in "Online Designer" or by using the "Check for Identifiers" page that is linked on the "Project Setup" page in the "Design your data collection instruments" box.

Examples

If a potential unmarked identifier is found, Checkmate will highlight the "Identifier?" column yellow and indicate the recommendation in Column S as seen below with the variable "dob". Note that age is an identifier for individuals older than 89 years old and should be marked as an identifier if the project will include that age group at any point during data collection.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Cntrl q to run macro															BP Sugges	BP Follow		BP Sugges	BP Follow
24	patient_document	demographics		file	Upload the patient's consent form															
25	first_name	demograp	Contact In	text	First Name						y									1
26	last_name	demographics		text	Last Name						y									1
27	address	demographics		notes	Street, City, State, ZIP						y									1
28	telephone_1	demographics		text	Phone number	Include Ar		phone			y									1,2
29	email	demographics		text	E-mail			email			y									1,2
30	dob	demographics		text	Date of birth			date_ymd											1	2
31	age	demographics		calc	Age (years rounddown(datediff([dob],'today','y'))														1	

Checkmate uses a conservative approach; therefore, some variables may be marked as potential identifiers even if they are not. This can be seen below with the variable "petname". Because the field label contains the word "name," the macro has suggested that this variable be marked as an identifier, even though it is not actually a HIPAA identifier.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Cntrl q to run macro															BP Sugges	BP Follow		BP Sugges	BP Follow
34	sex	demographics		radio	Sex	0, Female	1, Male													
35	petname	demographics		text	What is your pet's name?														1	
36	height2	baseline_data		text	Height (cm)			number											4,6	2,3
37	weight2	baseline_data		text	Weight (kilograms)			integer											4,6	2,3

If Checkmate identifies a field that could benefit from additional validation, such as adding minimum or maximum values for heights or weights, the data dictionary validation columns will be highlighted in yellow, and the corresponding recommendations will be numbered in Column S. In the above screenshot, the macro is recommending that the variables in rows 36 and 37 have minimums and maximums added.

91	General Recommendations:
92	5% of the fields in this study are free form text (Number Text fields: 34). If responses can be categorized, consider using a dropdown field type to reduce risk of data entry error and make the data easier to analyze.
93	0 out of 28 forms in this study have more than 30 fields.
94	Specific Recommendations (not all recommendations listed below will apply to this study):
95	1. Possible PHI. Consider placing a 'Y' in column K [Identifier?]
96	2. When using text fields, consider validating the field to expect a specific data type by entering it in column H [Text Validation Type]. Available options are: date, time, integer, number, zipcode, phone and email.
97	3. When entering common lab/physical exam values, consider validating the text field to expect number or integer values in column H [Text Validation Type]
98	4. For text fields validated as integer or number consider entering a minimum expected value in columns I [Text Validation Min] to decrease risk of data entry error.
99	5. For text fields validated as date consider entering a minimum value in columns I [Text Validation Min] to decrease risk of data entry error.
100	6. For text fields validated as integer or number consider entering a maximum expected value in columns J [Text Validation Max] to decrease risk of data entry error.
101	7. For text fields validated as date consider entering a maximum expected value in column J [Text Validation Max] to decrease risk of data entry error.
102	Fields in green meet best practice guidelines
103	Note: Age is an Identifier only if over 89

ILLINOIS
IHSI | Interdisciplinary Health
Sciences Institute