

REDCap Recruitment Tracker R Shiny Application - User Manual

CCTS Biostatistics Core

November 26, 2022

Contents

Purpose	1
Module 1. Verification & Setting	2
Instructions	3
Example	5
Module 2. Project Summary	5
Module 3. Recruitment Tracking	6
Tab 1. Screened Data	6
Tab 2. Eligible Data	10
Tab 3. Randomized Data	11
Module 4. Descriptive Statistics	13
Tab 1. Univariate Analysis	13
Tab 2. Bivariate Analysis	15

Purpose

REDCap (Research Electronic Data Capture) is a nationwide online survey and database. To visualize any project involving a randomized controlled trial (RCT) in REDCap, researchers can use the R Shiny application called Recruitment Tracker. Recruitment Tracker aims to establish a free, interactive, and web-based dashboard (uicccts.shinyapps.io/REDCap_Tracking). The platform R Shiny allows users to interact with and visualize the data according to their needs. The main function of this dashboard is to track the recruitment progress of an RCT project using tools and functionalities that provide real-time statistical results. Thus, this dashboard can support users in making decisions about recruitment progress and data collection based on the tracking reports.

The secondary function of the Recruitment Tracker is to visualize the characteristics of the participants assigned to receive randomization in RCTs. Researchers can perform univariate and bivariate analyses on all demographic and clinical variables of interest. All the results in the Recruitment Tracker presented in graphs and tables are able to be downloaded.

The workflow of the modules of this interactive online dashboard is illustrated in Figure 1. Recruitment Tracker displays the dataset in four modules of interest for users: “**Verification & Setting**,” “**Project Summary**,” “**Recruitment Tracking**,” and “**Descriptive Statistics**.” Recruitment Tracker was made publicly available in December 2022. Advanced R users may request the R Shiny code to adapt the code for their specific visualization of R Shiny apps.

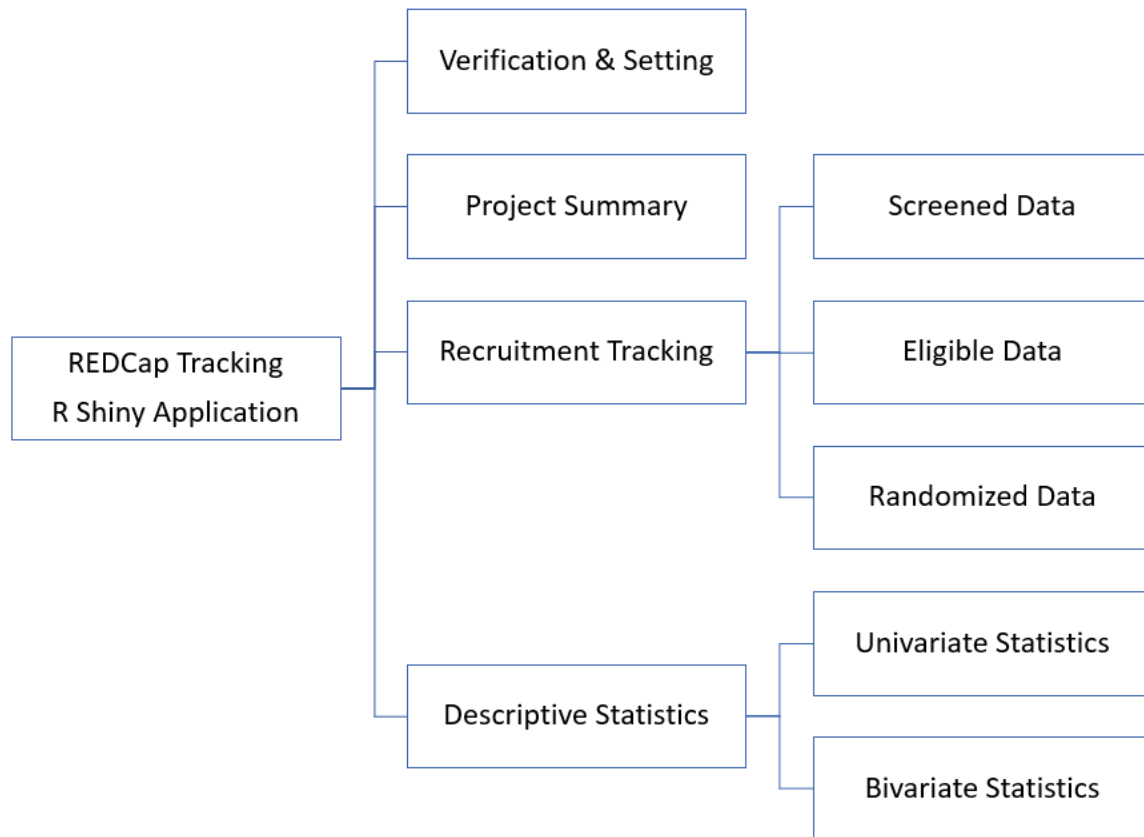


Figure 1: Workflow and modules of the interactive web-based dashboard for REDCap Recruitment Tracker R Shiny App

Module 1. Verification & Setting

This module is designed to import the user's long-form longitudinal dataset involving an RCT in REDCap (Figure 2 demonstrates the long-form format of an example dataset). Also, users must provide recruitment goals to compare with recruitment progress.

	A	B	C	D	E	F	G	H	I
1	study_id	redcap_event_name	screen_date	is_eligible	is_enrolled	exclusion_reason	enrolled_date	site	gender
2	1	baseline_arm_1	10/14/2019	1	1		10/15/2019	4	1
3	1	month_3_arm_1							
4	1	month_6_arm_1							
5	1	month_9_arm_1							
6	1	month_12_arm_1							
7	2	baseline_arm_1	5/10/2019	0				3	0
8	2	month_3_arm_1							
9	2	month_6_arm_1							
10	2	month_9_arm_1							
11	2	month_12_arm_1							
12	3	baseline_arm_1	11/6/2017	0				1	0
13	3	month_3_arm_1							
14	3	month_6_arm_1							
15	3	month_9_arm_1							
16	3	month_12_arm_1							
17	4	baseline_arm_1	11/21/2018	0				2	1
18	4	month_3_arm_1							
19	4	month_6_arm_1							
20	4	month_9_arm_1							

ClinicalTrialRecruit DATA 2022- (+)

Figure 2: An example of a long-form dataset

Instructions

Step 1 To import the dataset, users need to provide the token of the REDCap project of interest. Please note that it will appear as asterisks to ensure data confidentiality as you enter the token.

Then, the real-time dataset can be pulled from the online REDCap database directly. This dashboard will not import or analyze any variable involving personal information (i.e., identifier) to protect data privacy. Once the data are imported, variables are automatically separated into categorical variables and numerical variables—based on the data dictionary (codebook)—for the purpose of visualization and analyses.

Step 2 To produce the tracking statistics of the RCT project, users need to enter the following information into the boxes:

- Research Period of this Clinical Trial Project
- Total Targeted Number of Screened Participants
- Expected Rate of Eligibility
- Expected Rate of Randomization

Figure 3: Import the REDCap dataset via project token and then provide the initial setting information

Step 3 To identify the RCT selection process variables in REDCap, users need to provide the five variable names assigned to the information below. This app will present the recruitment progress based on the results of these five variables.

1. The date that the participant received the screening process (following the code rule: yyyy-mm-dd)
2. Participant eligibility (following the code rule: 1 = Yes and 0 = No)
3. Exclusion reason for the RCT project (no code rule)
4. Whether the participant receives the randomization (following the code rule: 1 = Yes and 0 = No)
5. The date that the participant received the randomization (following the code rule: yyyy-mm-dd)

Step 4 (Optional) Users can upload the variable list in the Variable List (Optional) box to observe specific variables of interest rather than all variables in the dataset. The content of the file should follow the rules below:

1. All the variable names match those in REDCap, including both spelling and capitalization. Please note that the variable list does not need to include the five variable names mentioned in Step 3.
2. All the variable names are listed in column 1 of the file (see Figure 4).
3. Rename “Sheet 1” on the Excel file to “variables.”
4. Please save the Excel file as a CSV file format.

Once the file is successfully uploaded, the message below the Import button will show “Uploaded file: filename.csv.”

	A
1	site
2	gender
3	race
4	age
5	bmi
6	treatment
7	
8	
9	
	variables

Figure 4: The data format of the optional variable list file (CSV file only)

Step 5 Once all the information is given, please press the green button labeled `Search`. If the message below the `Search` button shows “Dataset Successfully Imported,” it indicates that the dataset was imported from the REDCap database correctly, and the statistical results are ready to be revealed in the dashboard. However, the message will show “Fail to find the REDCap Data” if users provide the wrong RCT project token in REDCap or the wrong data format rule of the uploaded variable list. Then, users need to follow the above rules and go back to Step 1.

Step 6 (Optional) To save time for the future use of this dashboard, users can store all the setting information in Steps 1 to Step 4. After the message “Dataset Successfully Imported” is shown on the window, please press the green button labeled `Export` located below “Setting (Optional).” Then, all the setting information (including the optional variable list) will be exported into a downloaded JSON file to the user’s local desktop. Next time, when users want to use Recruitment Tracker, they can click the green button labeled `Import` and upload the JSON file. After doing this, users can press the green bottom labeled `Search` and import the data directly.

Example

This example demonstrates how to use this dashboard’s module of “Verification & Setting.” (see Figure 5). First, we provide the token A0C6E1AA3B6FF2F1EBF4A184FABE4378, of an artificial project we created in the REDCap database. The research period of this project is from 2017-07-01 to 2020-06-30. The total targeted number of screened participants is 2450, with an expected eligibility rate of 0.25 and an expected randomization rate of 0.90. In other words, in this RCT project, the total expected number of the randomized participants is 551.25 ($= 2450 \times 0.25 \times 0.90$). We randomly assigned the participants to receive either standard treatment or new treatment in this project.

The variable names for screening date, participant eligibility, the exclusion reason, randomization, and randomization date are `screen_date`, `is_eligible`, `exclusion_reason`, `is_enrolled`, and `enrolled_date`, respectively. Since we prefer to analyze a specific variable list, we upload a CSV file named “variable.csv.” Once we click the button of `Search`, the message below `Search` should show “Dataset Successfully Imported.” Then, we can further observe the statistical results in the “Project Summary,” “Recruitment Tracking,” and “Descriptive Statistics” modules.

Figure 5: Import the REDCap example dataset via token and provide the initial setting information

Module 2. Project Summary

After the data is successfully imported, this module provides the summary report of the RCT project. Figure 6 shows the summary of the example dataset. The research period of this RCT project is from 2017-07-01 to 2020-06-30. The total number of screened participants is 2450 in a total of 4 sites. Among the screened

participants, 650 were eligible for the RCT project. Among the eligible participants, there are 75 eligible participants excluded from the randomization step. The exclusion reasons include “Not meeting the inclusion criteria” (38), “Declined to participate” (20), and “Other reasons” (17).

Recruitment Tracker		UIC Center for Clinical and Translational Science	
Verification & Setting		Project Summary	
Project Summary			
Recruitment Tracking			
Descriptive Statistics			
User Manual			
Contact us			
Index		Content	
Research Period		2017-07-01 to 2020-06-30	
Observation Period		2017-07-01 to 2020-06-30	
The Total Number of Screened Participants		2450	
The Total Number of Eligible Participants		650	
The Total Number of Excluded Participants		75	
- Not meeting inclusion criteria		38	
- Declined to participate		20	
- Other Reasons		17	
The Total Number of Targeted Randomized Participants		551	
The Total Number of Actual Randomized Participants		575	
- Allocated to Standard Treatment		269	
- Allocated to New Treatment		306	
Randomization Accrual Rate		104.31%	
The Total number of Sites		4	
The Average Number of Randomized Participants per Site		143.75	
The Average Number Randomized Participants per Site per Month		4.64	
The Mean Number of Days from Screening to Randomization		15.26	
The Median Number of Days from Screening to Randomization		15	

Figure 6: Project Summary of the example dataset

Among the 575 participants ($= 650 - 75$) who received randomization, 269 received the standard treatment while 306 received the new treatment. The randomization accrual rate is 104.31% ($= \text{actual } 575 / \text{target } 551.25$). The average number of randomized participants per site is 143.75 ($= 575 / 4$). The mean number of days from screening to randomization is 15.26, and the median is 15.

Module 3. Recruitment Tracking

This module aims to track the screening, eligibility, and randomization progress. In the input panel, users need to decide the observation period, the unit of time (week/quarter/month/year), the stratified variable, and the substratified variable. Then, the corresponding plot will be revealed in the output panel. Users can utilize the green buttons labeled HTML, PNG, and EXCEL to download the interactive plot, static plot, and tracking report to the local desktop.

Tab 1. Screened Data

The tab “Screened Data” helps users visualize the real-time tracking statistics of the participants who received the screening process in the RCT project. For example, suppose users decide to observe the overall weekly

statistics without stratifying any variables during the observation period from 2017-07-01 to 2020-06-30. In that case, the plot will be as demonstrated in Figure 7. Users can interact with the plot by clicking any point on the line, and the statistics will be displayed simultaneously. Figure 7 shows that the cumulative number of screened participants in week 51, ending on 2018-06-22, is 973. If users click on EXCEL, an Excel file that includes the weekly tracking report is downloaded to the local desktop (see Figure 8).

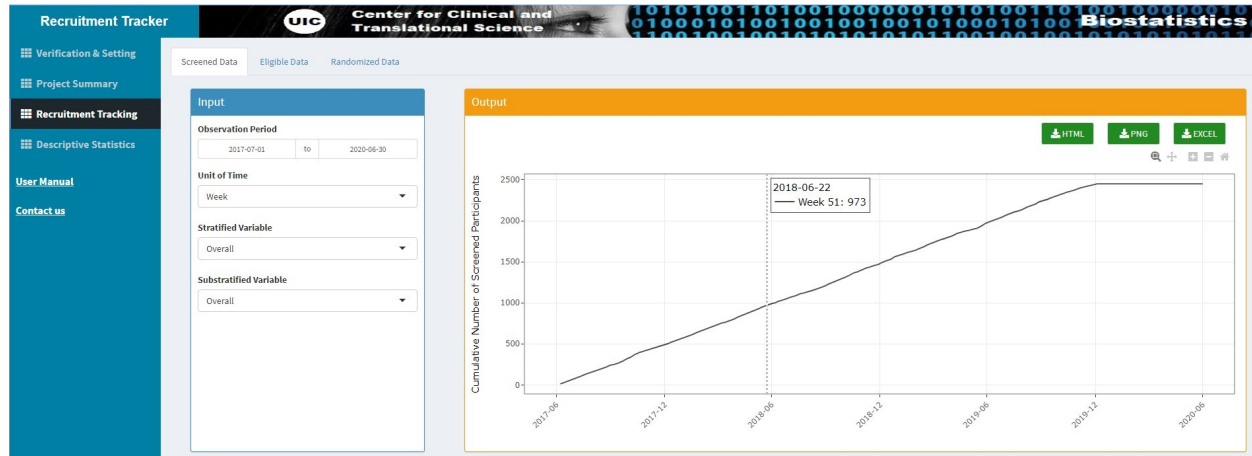


Figure 7: Tracking Statistics - Plot - Screened Weekly Data

	A	B	C	D	E
1	Research Period		2017-07-01 to 2020-06-30		
2	Observation Period		2017-07-01 to 2020-06-30		
3					
4	Participants Screened Overtime				
5	Week	Period		Overall	
6		Start Date	End Date	Actual	Cumulative Actual
7	Week 1	2017-07-01	2017-07-07	13	13
8	Week 2	2017-07-08	2017-07-14	18	31
9	Week 3	2017-07-15	2017-07-21	18	49
10	Week 4	2017-07-22	2017-07-28	19	68
11	Week 5	2017-07-29	2017-08-04	16	84
12	Week 6	2017-08-05	2017-08-11	23	107
13	Week 7	2017-08-12	2017-08-18	23	130
14	Week 8	2017-08-19	2017-08-25	18	148
15	Week 9	2017-08-26	2017-09-01	13	161
16	Week 10	2017-09-02	2017-09-08	19	180
17	Week 11	2017-09-09	2017-09-15	16	196
18	Week 12	2017-09-16	2017-09-22	21	217
19	Week 13	2017-09-23	2017-09-29	23	240
20	Week 14	2017-09-30	2017-10-06	10	250
21	Week 15	2017-10-07	2017-10-13	17	267
22	Week 16	2017-10-14	2017-10-20	21	288
23	Week 17	2017-10-21	2017-10-27	29	317
24	Week 18	2017-10-28	2017-11-03	21	338
25	Week 19	2017-11-04	2017-11-10	32	370
26	Week 20	2017-11-11	2017-11-17	24	394
27	Week 21	2017-11-18	2017-11-24	15	409
28	Week 22	2017-11-25	2017-12-01	16	425
29	Week 23	2017-12-02	2017-12-08	11	436
30	Week 24	2017-12-09	2017-12-15	18	454
31	Week 25	2017-12-16	2017-12-22	16	470
32	Overall				

Figure 8: Tracking Statistics - Table - Screened Weekly Data

If users decide to observe the weekly statistics by stratifying the variable gender during the observation period from 2017-07-01 to 2020-06-30, the interactive plot is shown in Figure 9. At the end of week 51 (2018-06-22), the cumulative number of participants who received a screening at sites Chicago, New York, Houston, and Los Angeles were 229, 238, 231, and 275, respectively. Again, the users can also obtain a weekly report stratified by the variable site when clicking on the EXCEL button. In the worksheet "Summary," the number of total screened participants stratified by the site is shown in Figure 10. In the worksheet "Stratified," the weekly report is stratified by the site (see Figure 11).

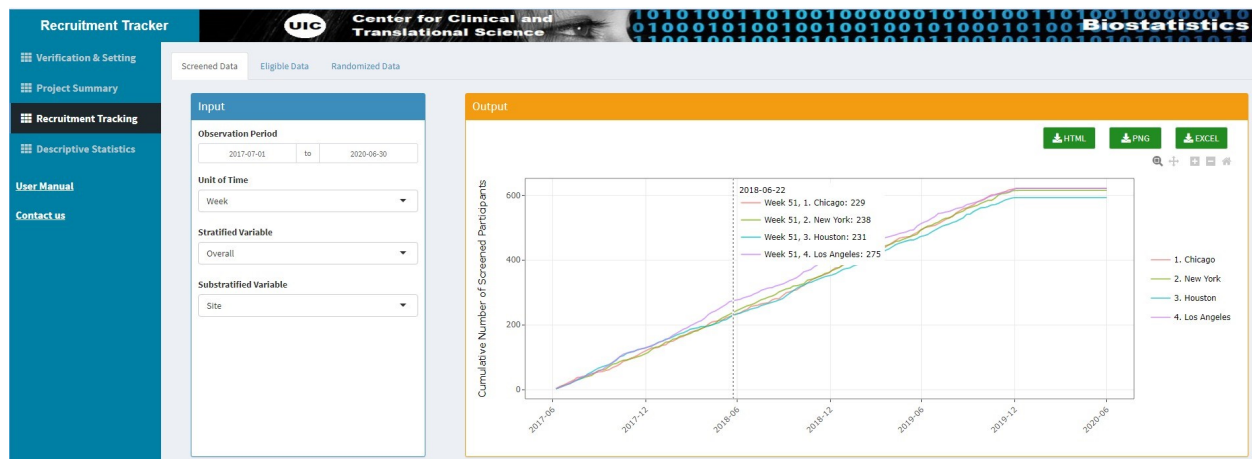


Figure 9: Tracking Statistics - Plot - Screened Weekly Data Stratified by Site

	A	B	C
1	Summary	Site	
2	1. Chicago	621	
3	2. New York	615	
4	3. Houston	593	
5	4. Los Angeles	621	
6	Total	2450	
7			
8			
9			
	Summary	Stratified	(+)

Figure 10: Tracking Statistics - Summary Table - Screened Data Stratified by Site

	A	B	C	D	E	F	G	H	I	J	K
1	Research Period		2017-07-01 to 2020-06-30								
2	Observation Period		2017-07-01 to 2020-06-30								
3											
4	Participants Screened Overtime										
5	Week	Period		Site: 1. Chicago		Site: 2. New York		Site: 3. Houston		Site: 4. Los Angeles	
6		Start Date	End Date	Actual	Cumulative Actual	Actual	Cumulative Actual	Actual	Cumulative Actual	Actual	Cumulative Actual
7	Week 1	2017-07-01	2017-07-07	3	3	3	3	3	3	4	4
8	Week 2	2017-07-08	2017-07-14	7	10	4	7	4	7	3	7
9	Week 3	2017-07-15	2017-07-21	4	14	6	13	4	11	4	11
10	Week 4	2017-07-22	2017-07-28	6	20	3	16	5	16	5	16
11	Week 5	2017-07-29	2017-08-04	5	25	3	19	4	20	4	20
12	Week 6	2017-08-05	2017-08-11	5	30	8	27	5	25	5	25
13	Week 7	2017-08-12	2017-08-18	8	38	3	30	5	30	7	32
14	Week 8	2017-08-19	2017-08-25	1	39	4	34	6	36	7	39
15	Week 9	2017-08-26	2017-09-01	3	42	4	38	4	40	2	41
16	Week 10	2017-09-02	2017-09-08	4	46	3	41	9	49	3	44
17	Week 11	2017-09-09	2017-09-15	3	49	3	44	5	54	5	49
18	Week 12	2017-09-16	2017-09-22	2	51	8	52	7	61	4	53
19	Week 13	2017-09-23	2017-09-29	4	55	7	59	6	67	6	59
20	Week 14	2017-09-30	2017-10-06	1	56	2	61	4	71	3	62
21	Week 15	2017-10-07	2017-10-13	3	59	3	64	3	74	8	70
22	Week 16	2017-10-14	2017-10-20	3	62	8	72	4	78	6	76
23	Week 17	2017-10-21	2017-10-27	7	69	7	79	6	84	9	85
24	Week 18	2017-10-28	2017-11-03	4	73	2	81	7	91	8	93
25	Week 19	2017-11-04	2017-11-10	6	79	6	87	11	102	9	102
26	Week 20	2017-11-11	2017-11-17	9	88	4	91	3	105	8	110
27	Week 21	2017-11-18	2017-11-24	3	91	1	92	8	113	3	113
28	Week 22	2017-11-25	2017-12-01	6	97	3	95	3	116	4	117
29	Week 23	2017-12-02	2017-12-08	4	101	3	98	2	118	2	119
30	Week 24	2017-12-09	2017-12-15	5	106	3	101	6	124	4	123
	Summary		Stratified								

Figure 11: Tracking Statistics - Stratified Table - Screened Weekly Data Stratified by Site

Tab 2. Eligible Data

The tab “Eligible Data” helps users visualize the real-time tracking statistics of the eligible participants in the RCT project. Similar to the operation of the “Screened Data” tab, if users decide to observe the monthly statistics by stratifying the variable gender during the observation period from 2017-07-01 to 2017-12-31, the interactive plot is shown in Figure 12. Users can interact with the plot to further monitor the monthly eligibility of different genders. For instance, until the end of month 3 (ending on 2017-09-30), the cumulative number of male participants eligible for this RCT project is 48. Again, the users can also obtain a report stratified by the variable gender when clicking on the EXCEL button (see Figure 13).



Figure 12: Tracking Statistics - Plot - Eligible Monthly Data Stratified by the Variable Gender

	A	B	C	D	E
1	Research Period		2017-07-01 to 2020-06-30		
2	Observation Period		2017-07-01 to 2017-12-31		
3					
4	Participants Eligible Overtime by Gender				
5	Month	Period		Overall	
6		Start Date	End Date	Actual	Cumulative Actual
7	Month 1	2017-07-01	2017-07-31	16	16
8	Month 2	2017-08-01	2017-08-31	14	30
9	Month 3	2017-09-01	2017-09-30	18	48
10	Month 4	2017-10-01	2017-10-31	21	69
11	Month 5	2017-11-01	2017-11-30	14	83
12	Month 6	2017-12-01	2017-12-31	15	98
13	Month 7	2018-01-01	2018-01-31		
14	Month 8	2018-02-01	2018-02-28		
15	Month 9	2018-03-01	2018-03-31		
16	Month 10	2018-04-01	2018-04-30		
17	Month 11	2018-05-01	2018-05-31		
18	Month 12	2018-06-01	2018-06-30		
19	Month 13	2018-07-01	2018-07-31		
20	Month 14	2018-08-01	2018-08-31		
21	Month 15	2018-09-01	2018-09-30		
22	Month 16	2018-10-01	2018-10-31		
23	Month 17	2018-11-01	2018-11-30		
24	Month 18	2018-12-01	2018-12-31		
<div> <div>0. Male</div> <div>1. Female</div> <div>+</div> </div>					

Figure 13: Tracking Statistics - Table - Eligible Monthly Data stratified by the Variable Gender

Tab 3. Randomized Data

The tab “Randomized Data” helps users visualize the real-time tracking statistics of the participants who received the randomization in the RCT project. Similar to the operation of the “Screened Data” tab, the interactive plot is shown if users decide to observe the monthly statistics by stratifying the variables site and gender during the observation period from 2017-07-01 to 2020-06-30 (see Figure 14). Users can interact with the plot to further monitor the monthly randomization at different sites and genders. For instance, until the end of month 19 (ending on 2019-01-31), the cumulative target and the cumulative actual number of female participants who received the randomization at site Houston were 36.1 and 58, respectively. Again, the users can also obtain a report stratified by the variable site and gender when clicking on the EXCEL button (see Figures 15-16).

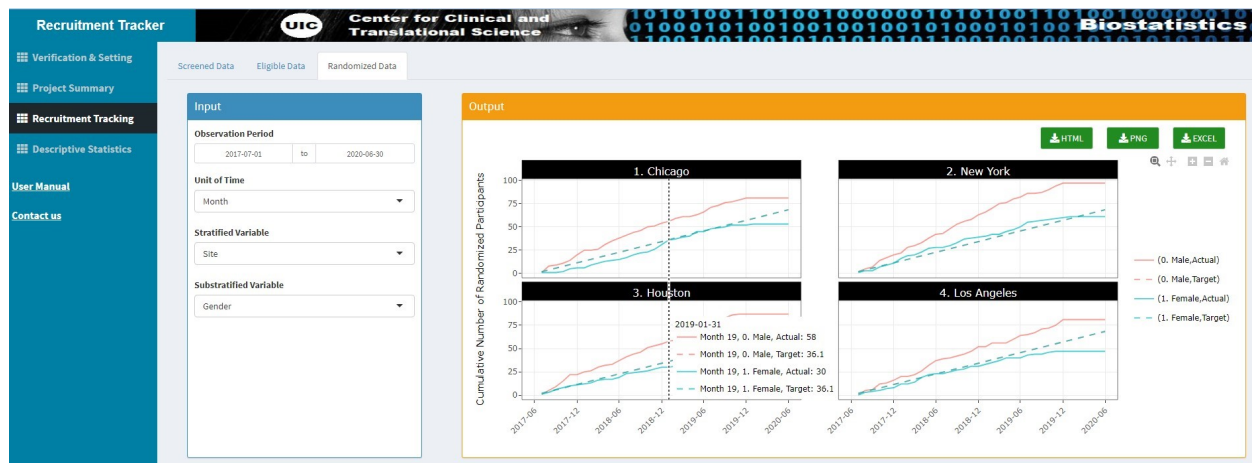


Figure 14: Tracking Statistics - Plot - Randomized Monthly Data Stratified by the Variable Site and Gender

	A	B	C	D	E	F
1	Summary		Gender			
2			0. Male	1. Female	Total	
3	Site	1. Chicago	81	53	134	
4		2. New York	97	61	158	
5		3. Houston	87	50	137	
6		4. Los Angeles	81	47	128	
7		Total	346	211	557	
8						
9						
		Summary	1. Chicago	2. New York	3. Houston	4. Los Angeles

Figure 15: Tracking Statistics - Summary Table - Randomized Data stratified by the Variable Site and Gender

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Research Period	2017-07-01 to 2020-06-30													
2	Observation Period	2017-07-01 to 2020-06-30													
3	Participants Randomized Overtime by Site														
4	Gender: 0. Male														
5	Month	Start Date	End Date	Target	Actual	Accrual Rate	Cumulative Target	Cumulative Actual	Cumulative Accrual Rate	Target	Actual	Accrual Rate	Cumulative Target	Cumulative Actual	Cumulative Accrual Rate
6	Month 1	2017-07-01	2017-07-31	1.9	1	53%	1.9	1	53%	1.9	1	53%	1.9	1	53%
7	Month 2	2017-08-01	2017-08-31	1.9	4	211%	3.8	5	132%	1.9	2	105%	3.8	3	79%
8	Month 3	2017-09-01	2017-09-30	1.9	4	211%	5.7	9	158%	1.9	3	158%	5.7	6	105%
9	Month 4	2017-10-01	2017-10-31	1.9	6	316%	7.6	15	197%	1.9	2	105%	7.6	8	105%
10	Month 5	2017-11-01	2017-11-30	1.9	7	368%	9.5	22	232%	1.9	2	105%	9.5	10	105%
11	Month 6	2017-12-01	2017-12-31	1.9	0	0%	11.4	22	193%	1.9	1	53%	11.4	11	96%
12	Month 7	2018-01-01	2018-01-31	1.9	3	158%	13.3	25	188%	1.9	1	53%	13.3	12	90%
13	Month 8	2018-02-01	2018-02-28	1.9	1	53%	15.2	26	171%	1.9	1	53%	15.2	13	86%
14	Month 9	2018-03-01	2018-03-31	1.9	4	211%	17.1	30	175%	1.9	3	158%	17.1	16	94%
15	Month 10	2018-04-01	2018-04-30	1.9	2	105%	19.0	32	168%	1.9	1	53%	19.0	17	89%
16	Month 11	2018-05-01	2018-05-31	1.9	1	53%	20.9	33	158%	1.9	0	0%	20.9	17	81%
17	Month 12	2018-06-01	2018-06-30	1.9	4	211%	22.8	37	162%	1.9	2	105%	22.8	19	83%
18	Month 13	2018-07-01	2018-07-31	1.9	4	211%	24.7	41	166%	1.9	4	211%	24.7	23	93%
19	Month 14	2018-08-01	2018-08-31	1.9	3	158%	26.6	44	165%	1.9	1	53%	26.6	24	90%
20	Month 15	2018-09-01	2018-09-30	1.9	2	105%	28.5	46	161%	1.9	1	53%	28.5	25	88%
21	Month 16	2018-10-01	2018-10-31	1.9	5	263%	30.4	51	168%	1.9	1	53%	30.4	26	86%
22	Month 17	2018-11-01	2018-11-30	1.9	2	105%	32.3	53	164%	1.9	2	105%	32.3	28	87%
23	Month 18	2018-12-01	2018-12-31	1.9	2	105%	34.2	55	161%	1.9	2	105%	34.2	30	88%
24	Month 19	2019-01-01	2019-01-31	1.9	3	158%	36.1	58	161%	1.9	0	0%	36.1	30	83%
25	Month 20	2019-02-01	2019-02-28	1.9	1	53%	38.0	59	155%	1.9	1	53%	38.0	31	82%
26	Month 21	2019-03-01	2019-03-31	1.9	5	263%	39.9	64	160%	1.9	3	158%	39.9	34	85%
27	Month 22	2019-04-01	2019-04-30	1.9	3	158%	41.8	67	160%	1.9	2	105%	41.8	36	86%
28	Month 23	2019-05-01	2019-05-31	1.9	3	158%	43.7	70	160%	1.9	4	211%	43.7	40	92%
29	Month 24	2019-06-01	2019-06-30	1.9	2	105%	45.6	72	158%	1.9	0	0%	45.6	40	88%
30	Month 25	2019-07-01	2019-07-31	1.9	7	368%	47.5	79	166%	1.9	1	53%	47.5	41	86%
31	Month 26	2019-08-01	2019-08-31	1.9	3	158%	49.4	82	166%	1.9	2	105%	49.4	43	87%
32	Month 27	2019-09-01	2019-09-30	1.9	1	53%	51.3	83	162%	1.9	0	0%	51.3	43	84%
33	Month 28	2019-10-01	2019-10-31	1.9	3	158%	53.2	86	162%	1.9	4	211%	53.2	47	88%
34	Month 29	2019-11-01	2019-11-30	1.9	1	53%	55.1	87	158%	1.9	0	0%	55.1	47	85%
35	Month 30	2019-12-01	2019-12-31	1.9	0	0%	57.0	87	153%	1.9	2	105%	57.0	49	86%
36	Month 31	2020-01-01	2020-01-31	1.9	0	0%	58.9	87	148%	1.9	1	53%	58.9	50	85%
37	Month 32	2020-02-01	2020-02-29	1.9	0	0%	60.8	87	143%	1.9	0	0%	60.8	50	82%
38	Month 33	2020-03-01	2020-03-31	1.9	0	0%	62.7	87	139%	1.9	0	0%	62.7	50	80%
39	Month 34	2020-04-01	2020-04-30	1.9	0	0%	64.6	87	135%	1.9	0	0%	64.6	50	77%
40	Month 35	2020-05-01	2020-05-31	1.9	0	0%	66.5	87	131%	1.9	0	0%	66.5	50	75%
41	Month 36	2020-06-01	2020-06-30	1.9	0	0%	68.4	87	127%	1.9	0	0%	68.4	50	73%
42															
43															
44															

Figure 16: Tracking Statistics - Table - Randomized Monthly Data stratified by the Variable Site and Gender

Module 4. Descriptive Statistics

This module aims to visualize the real-time characteristics of participants who received the treatment's randomization in the RCT project. Users can perform univariate and bivariate analyses on all the variables of interest. In the tab "Univariate Statistics," users need to decide the variables of interest to perform the analysis in the input panel. Users can choose multiple variables in the input panel and then obtain corresponding plots simultaneously in the output panel.

Tab 1. Univariate Analysis

Categorical Variable – Bar Chart The procedure to analyze categorical variables is as follows. Users must select the variables of interest in the Categorical Variables box in the input panel. Then the corresponding bar charts will be displayed in the output panel. When users click on one of the bars within the plot, the count and percentage will be revealed. The sample size for the total, valid, and missing participants is displayed under the title of the plot (i.e., the variable name). Users can utilize the green button labels HTML, PNG, and EXCEL to download the interactive plot, static plot, and frequency table to the local desktop.

For example, if the variable "Gender" is selected, then the interactive bar chart is shown in Figure 17. The title of the plot is "Gender," and the sample size for the total, valid, and missing participants are 575, 557, and 18, respectively. When users click on the bar "Male," it will show that the count (valid sample size) for the Male group is 346, and that percentage is 62.1% (= 346 / 557). If users click on the EXCEL button, an excel file that includes the frequency table for the variable "Gender" is downloaded to the local desktop (see Figure 18).

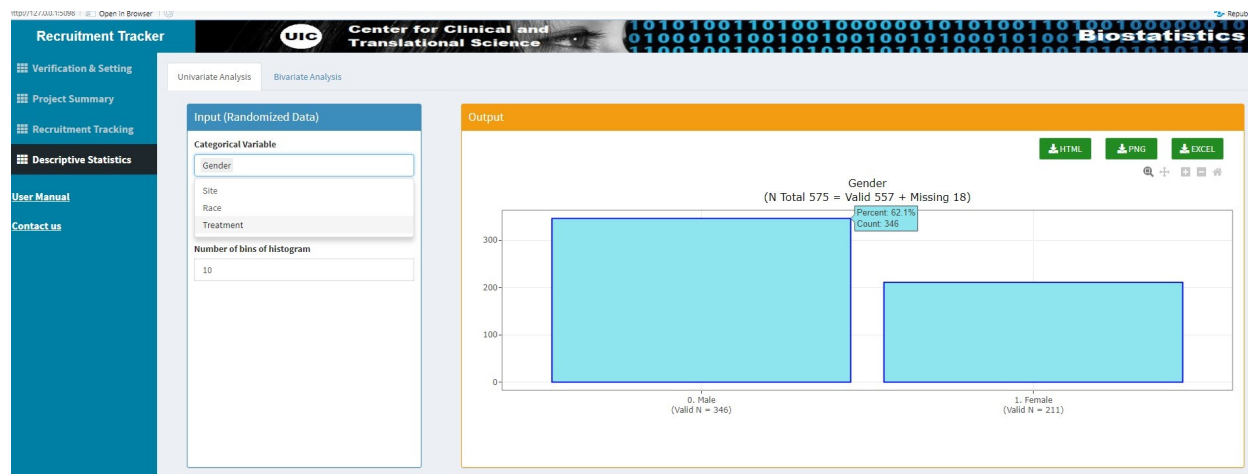


Figure 17: Univariate Statistics – Bar Chart for Categorical Variable 'Gender'

	A	B	C	D	E
1	Research Period	2017-07-01 to 2020-06-30			
2	Observation Period	2017-07-01 to 2020-06-30			
3					
4	Gender	Frequency	Percent	Cumulative Frequency	Cumulative Percent
5	0. Male	346	62.12	346	62.12
6	1. Female	211	37.88	557	100.00
7	Frequency Missing = 18				

Figure 18: Univariate Statistics – Frequency Table for Categorical Variable 'Gender'

Numerical Variable - Histogram The procedure to analyze numerical variables is as follows. Users must select the variables of interest in the box labeled **Numerical Variable** and decide the number of bins of the histogram (the default value is 10) in the input panel. The corresponding histogram will be displayed on the output panel. When users click on one of the bars within the plot, the bar's count and range will be revealed. The sample size and the summary statistics, including Mean/SD/Median/IQR, will be displayed under the title of the plot.

For example, if the variable "BMI" is selected, the interactive histogram is shown in Figure 19. The title of the plot is "BMI," and the sample size for the total, valid, and missing participants are 575, 575, and 0, respectively. The summary statistics of the variable BMI, including Mean, SD, Median, IQR, are 26.44, 3.55, 26.32, and 5.00, respectively. When users click on the bar between the group "30-32," it will show that the count of this group is 62. Users can also download the table to observe this BMI's minimum, first quartile (Q1), third quartile (Q3), and maximum values (see Figure 20).

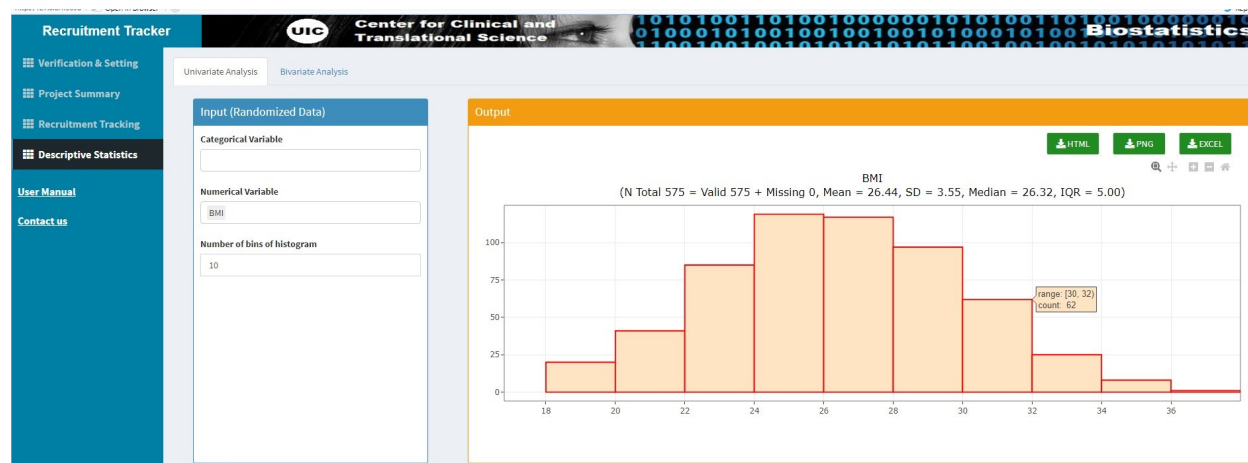


Figure 19: Univariate Statistics – Histogram for Numerical Variable 'BMI'

	A	B	C	D	E	F	G	H	I	J	K
1	Research Period		2017-07-01 to 2020-06-30								
2	Observation Period		2017-07-01 to 2020-06-30								
3											
4	Variable	N	N.Miss	Mean	Std.Dev	IQR	Minimum	Q1	Median	Q3	Maximum
5	BMI	575	0	26.44	3.55	5.00	18.50	23.95	26.32	28.95	37.52
6											

Figure 20: Univariate Statistics – Summary Statistics for Numerical Variable 'BMI'

Tab 2. Bivariate Analysis

In the “Bivariate Statistics” tab is the stratified results. Stratified statistics identify the relationship between two variables by breaking up the data into smaller groups, or strata, to look for differences between and within each stratum. Using this dashboard, users need to decide the primary and the stratified variables to perform the stratified analysis. The stratified variable will separate the primary variable (i.e., the fixed option) into groups. In the input panel, users first select the primary variable from the drop-down variable list, then choose one or multiple stratified variables (i.e., one or more options) to perform the stratification. Once these variables are decided, the corresponding plots will be simultaneously displayed in the output panel.

Primary - Categorical Variable vs. Stratified - Categorical Variable: Stratified Bar Chart The stratified bar chart will be displayed in the output panel if we would like to observe the relationship between two categorical variables. For example, when the primary variable “Treatment” and the stratified variable “Gender” are chosen, the interactive bar chart with the title “Treatment by Gender” is shown in Figure 21. The sample size for the total, valid, and missing participants is 575, 557, and 18. When users click on the bar “Female” within the group of “Standard Treatment,” it will show that the count of the female who received “Standard Treatment” is 167, with a percentage of 64.0% ($= 167 / 261$). If users click on the EXCEL button, the contingency table—which contains the stratified results—will be downloaded as an excel file (see Figure 22).

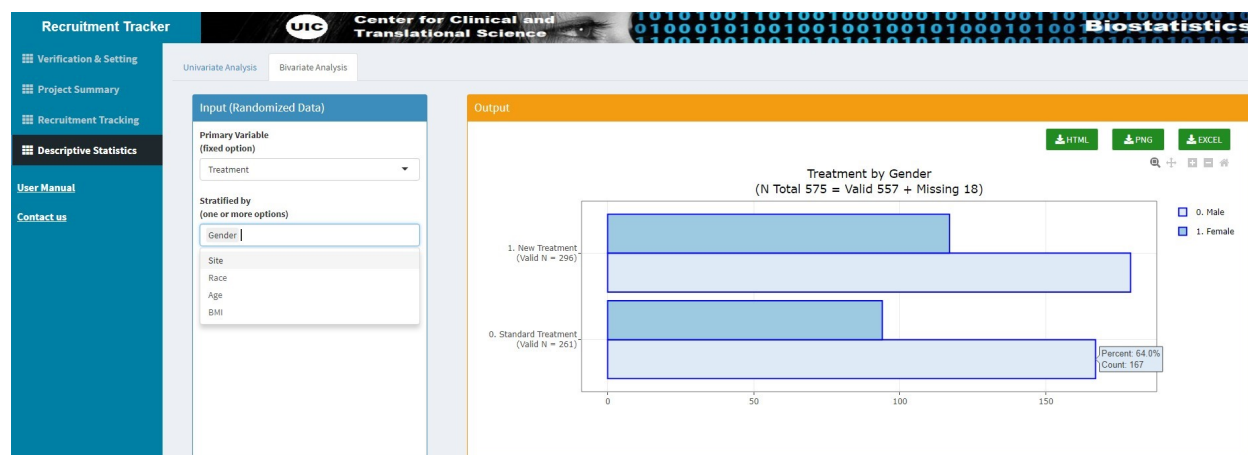


Figure 21: Bivariate Statistics – Bar Chart for the Variable 'Treatment' Stratified by 'Gender'

	A	B	C	D	E
1	Research Period		2017-07-01 to 2020-06-30		
2	Observation Period		2017-07-01 to 2020-06-30		
3					
4	Frequency		Treatment		
5			0. Standard Treatment	1. New Treatment	Total
6	Gender	0. Male	167	179	346
7		1. Female	94	117	211
8		Total	261	296	557
9	Column Percent		Treatment		
10			0. Standard Treatment	1. New Treatment	Total
11	Gender	0. Male	63.98	60.47	62.12
12		1. Female	36.02	39.53	37.88
13		Total	100.00	100.00	100.00
14	Row Percent		Treatment		
15			0. Standard Treatment	1. New Treatment	Total
16	Gender	0. Male	48.27	51.73	100.00
17		1. Female	44.55	55.45	100.00
18		Total	46.86	53.14	100.00
19	Frequency Missing = 18				
20					
21					

Figure 22: Bivariate Statistics – Contingency Table for the Variable 'Treatment' Stratified by 'Gender'

Primary - Categorical Variable vs. Stratified - Numerical Variable: Grouped Boxplot A boxplot is a standardized method of displaying the distribution of the data based on the summary statistics of a given variable, which include minimum (min), first quartile (q1), median (q2), third quartile (q3), and maximum (max). Also, this plot will indicate the value of outliers. A grouped boxplot is the visual representation of the difference in the subgroup distribution, which users can utilize to observe the relationship between a categorical and numerical variable.

For example, Figure 23 demonstrates that the variable "Age" is stratified by "Treatment." The grouped boxplot visualizes the distribution of two subgroups of "Treatment," including "Standard Treatment" and "New Treatment." The age distribution among the two types of treatments looks similar. To examine the age distribution of standard treatment, users may click on the green box "Standard Treatment," and then the

summary statistics will be shown in the plot. The median age among the participants who received standard treatment is 50, with min of 18 and max of 75. Again, the users can also obtain stratified summary statistics when clicking on the EXCEL button (see Figure 24).

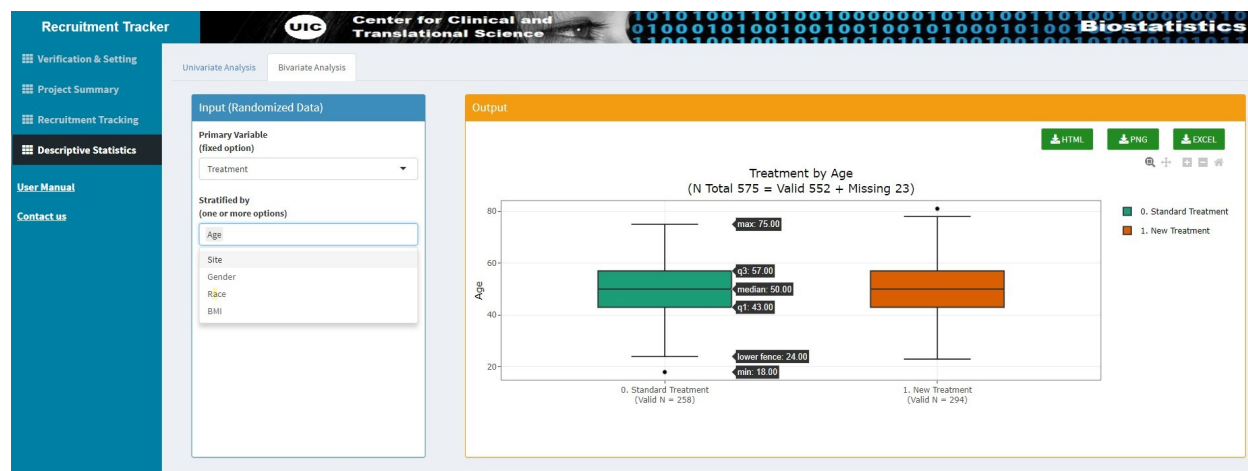


Figure 23: Bivariate Statistics – Grouped Boxplot for the Variable 'Age' Stratified by 'Treatment'

	A	B	C	D	E	F	G	H	I	J	K
1	Research Period		2017-07-01 to 2020-06-30								
2	Observation Period		2017-07-01 to 2020-06-30								
3											
4	Treatment = 0. Standard Treatment										
5	Variable	N	N.Miss	Mean	Std.Dev	IQR	Minimum	Q1	Median	Q3	Maximum
6	Age	258	11	49.93	9.88	13.75	18.00	43.25	50.00	57.00	75.00
7											
8	Treatment = 1. New Treatment										
9	Variable	N	N.Miss	Mean	Std.Dev	IQR	Minimum	Q1	Median	Q3	Maximum
10	Age	294	12	49.98	10.19	14.00	23.00	43.00	50.00	57.00	81.00

Figure 24: Bivariate Statistics – Stratified Summary Statistics for the Variable 'Age' stratified by 'Treatment'

Primary - Numerical Variable vs. Stratified - Numerical Variable: Scatter Plot The most straightforward stratified plot is an X-Y plot with the measured variable on the vertical axis and a variable describing the different strata on the horizontal axis. Thus, we will use the scatter plot to observe the relationship between two numerical variables. The scatter plot will be displayed in the output panel once the primary and stratified variables are chosen in the input panel. To describe the overall trend of the data, a polynomial model (the blue curve) with its 95% confidence interval (the gray area) is included in the plot. If users want to know the value of the paired data, they can click on any black point on the scatter plot, and the numerical value of the two variables will be displayed.

For example, when the primary variable "Age" and the stratified variable "BMI" are chosen, the interactive bar chart with the title "Treatment by BMI" is shown in Figure 25. The sample size for the total, valid, and missing participants is 575, 552, and 23. The primary variable, "Age," will be on the horizontal axis, and the stratified variable, "BMI," will be on the vertical axis. If users click the point in the lower-left corner, then the value of the point will be revealed (Age = 23, BMI = 22.80). Users can further download the table of stratified summary statistics to observe Pearson correlation and Spearman correlation by clicking the EXCEL button (see Figure 26).

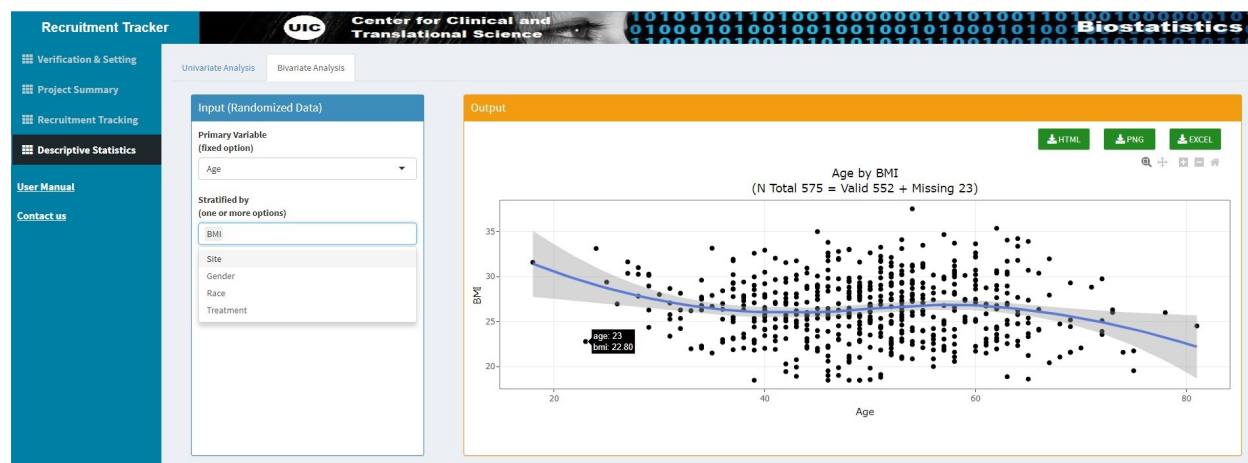


Figure 25: Bivariate Statistics – Scatter Plot for Numerical Variables 'Age' and 'BMI'

	A	B	C	D
1	Research Period		2017-07-01 to 2020-06-30	
2	Observation Period		2017-07-01 to 2020-06-30	
3				
4	Pearson Correlation Coefficients (N = 552)			
5		Age	BMI	
6	Age	1.0000	-0.0423	
7	BMI	-0.0423	1.0000	
8				
9	Spearman Correlation Coefficients (N = 552)			
10		Age	BMI	
11	Age	1.0000	-0.0224	
12	BMI	-0.0224	1.0000	
13				

Figure 26: Bivariate Statistics – Stratified Summary Statistics for Numerical Variables 'Age' and 'BMI'

Primary - Numerical Variable vs. Stratified - Categorical Variable: Grouped Boxplot If the primary variable is a numerical variable “Age” and the stratified variable is a categorical variable “Site”, then Figure 27 demonstrates that the variable “Age” is stratified by “Site.” The grouped boxplot visualizes the distribution of four subgroups of “Site,” including “Chicago,” “New York,” “Houston,” and “Los Angeles.” The age distribution among the four sites looks similar to each other. To examine the age distribution of the site Chicago, users may click on the green box “Chicago,” and then the summary statistics will be shown in the plot. The median age among Chicago participants is 51, with min of 26 and max of 75. Users can further download the table of stratified summary statistics by clicking the Table button (see Figure 28).

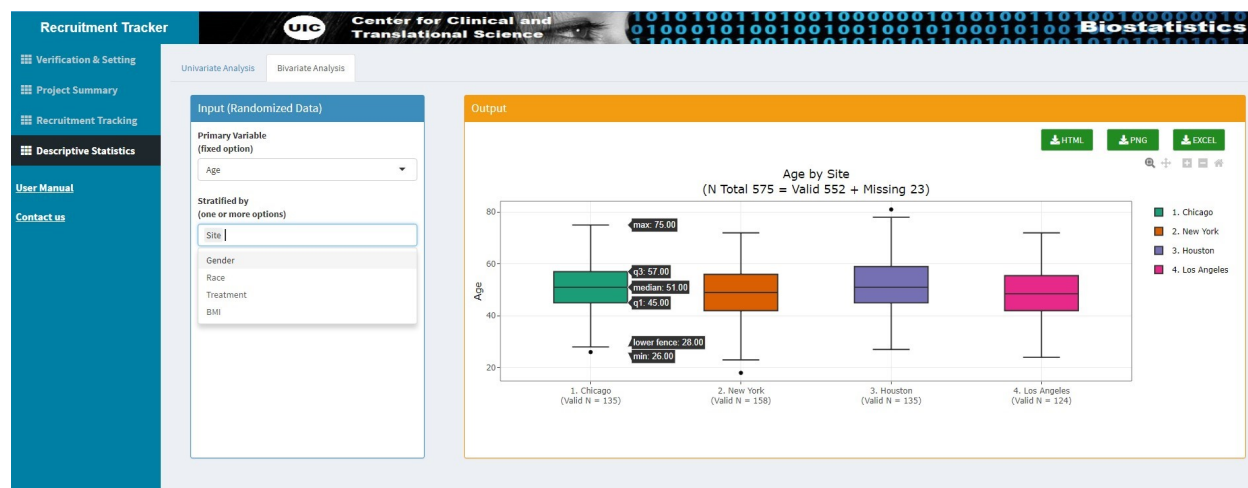


Figure 27: Bivariate Statistics – Grouped Boxplot for Variable 'Age' Stratified by 'Site'

	A	B	C	D	E	F	G	H	I	J	K	L
1	Research Period		2017-07-01 to 2020-06-30									
2	Observation Period		2017-07-01 to 2020-06-30									
3												
4												
5	Variable	Level	N	N Miss	Mean	Std Dev	IQR	Minimum	Q1	Median	Q3	Maximum
6	Overall		575	23	49.95	10.04	14.00	18.00	43.00	50.00	57.00	81.00
7	Site	1. Chicago	140	5	51.31	10.14	12.00	26.00	45.00	51.00	57.00	75.00
8		2. New York	160	2	48.84	9.97	14.00	18.00	42.00	49.00	56.00	72.00
9		3. Houston	142	7	51.06	10.09	14.00	27.00	45.00	51.00	59.00	81.00
10		4. Los Angeles	133	9	48.70	9.74	13.25	24.00	42.00	48.50	55.25	72.00

Figure 28: Bivariate Statistics – Stratified Summary Statistics for the Variable 'Age' stratified by 'Site'