Automated Generation of CONSORT Diagrams Using Relational Database Software

Teresa O'Leary[1,2]   June Weiss[1,2]   Benjamin Toll[3]   Cynthia Brandt[1,2,4]   Steven L. Bernstein[1,2,5]

[1] Department of Emergency Medicine, Yale School of Medicine, New Haven, Connecticut
[2] Yale Center for Implementation Science, Yale School of Medicine, New Haven, Connecticut
[3] Department of Public Health Sciences, Medical University of South Carolina, Charleston, South Carolina
[4] Department of Anesthesiology, Yale School of Medicine, New Haven, Connecticut
[5] Yale Comprehensive Cancer Center, Yale School of Medicine, New Haven, Connecticut

Address for correspondence Steven L. Bernstein, MD, Department of Emergency Medicine, Yale Center for Implementation Science, Yale School of Medicine, 464 Congress Avenue, Suite 260, New Haven, CT 06519, United States (e-mail: steven.bernstein@yale.edu).

Abstract

Background   Investigators conducting prospective clinical trials must report patient flow using the Consolidated Standards of Reporting Trials (CONSORT) statement. Depending on how data are collected, this can be a laborious, time-intensive process. However, because many trials enter data electronically, CONSORT diagrams may be generated in an automated fashion.

Objective   Our objective was to use an off-the-shelf software to develop a technique to generate CONSORT diagrams automatically.

Methods   During a recent trial, data were entered into FileMaker Pro, a commercially available software, at enrollment and three waves of follow-up. Patient-level data were coded to automatically generate CONSORT diagrams for use by the study team.

Results   From August 2012 to July 2014, 1,044 participants were enrolled. CONSORT diagrams were generated weekly for study team meetings to track follow-ups at 1, 6, and 12 months, for 960 (92%), 921 (90%), and 871 (88%) participants who were contacted or deceased, respectively. Reasons for loss to follow-up were captured at each follow-up.

Conclusion   CONSORT diagrams can be generated using a standard software for any trial and can facilitate data collection, project management, and reporting.

Keywords
► clinical research
► data management
► relational database

Background and Significance

Data management is a critical component in conducting clinical trials. Accurate and complete data capture is essential to minimize the risk of bias, allow for the assessment of fidelity to protocol, minimize loss of participants to follow-up, and identify potential adverse events.[1] Clinical Data Management Systems (CDMSs) can help collect, store, and manage data, and come in various levels of complexity that range from enterprise-level systems, such as OnCore,[2] to research-oriented systems, such as REDCap.[3] For smaller studies, simple off-the-shelf software equipped with graphical user interface tools such as Microsoft Access or FileMaker Pro can be used to create data collection instruments and organize trial data in a more secure way than paper or flat files.[4] Most of these systems also have tools that can facilitate the management of trial data, such as creating case report forms directly on an electronic device (tablet, computer,

smartphone), thus eliminating the need to enter data from paper.[5] While CDMSs are not yet ubiquitous, the adoption of such systems can increase productivity and overall data accuracy and can create superior project management tools.[6] Computational software capitalizes on pattern recognition of both project management data and human subject data that can yield significant benefits of time and resource management to the research team.[6]

Publication of randomized clinical trials requires submission of a Consolidated Standards of Reporting Trial (CONSORT) diagram that illustrates the flow of participants through all phases of the trial. It is a ubiquitous feature of reports of prospective randomized trials.[5] The CONSORT framework allows others to assess the generalizability and validity of trial results.[5] The diagram provides a statistical summary of study data, including screening for eligibility, enrollment, losses to follow-up, and exclusions during analysis.[5]

For trials with many participants, standard methods to create the CONSORT diagram may be time- and resource-intensive, requiring staff to collect screening data on paper to later be entered into flat files.

We conducted a search of the English-language literature, using Medline, from 1996 to present. (The CONSORT statement was first published in 1996.) The term "*CONSORT statement*" was cross-referenced with "automated," "automation," "building," "construction," "generation," "database," "database management," "database management system," and "clinical database management system"; 55 articles were found and reviewed by two study authors. A separate search, using Medline, from 1966 to present, used the term *database management systems* and cross-referenced it with "randomized controlled trials"; 31 articles were found. No articles were found that discussed the use of relational database software to generate CONSORT flow diagrams automatically.

Objective

Using software tools and information gathered during electronic data collection, we provide an example of how to automatically generate a CONSORT flow diagram at any time during the trial.

Using FileMaker Pro, we developed data collection tools and a method to generate CONSORT diagrams automatically. We provide, as an example, a diagram from a trial recently completed by our group.[7] We found that this method yielded significant benefits in accuracy, staff use, and follow-up management.

Methods

We developed automated CONSORT diagrams for a clinical trial of tobacco dependence treatment. The trial, conducted in an academic medical center in the northeastern United States, randomized physicians to control or intervention arms. Intervention physicians viewed a decision prompt when logging on to the electronic medical record (EMR) of an adult smoker recently admitted to the hospital. The prompt offered to take the physician to an order set that included medication prescription, referral to a state smokers' quitline, autopopulation of the problem list with "tobacco use disorder," and e-mail notification of the patient's primary care provider. Details of the parent study have been reported previously.[7]

FileMaker Pro (Apple Inc., Cupertino, California, United States), an off-the-shelf relational database application, was used to collect the clinical trial data and generate the CONSORT diagram (►Fig. 1).[8] Data were extracted from the following three tables:

- Initial screening and eligibility data.
- Follow-up data for participants at 1, 6, and 12 months postenrollment.
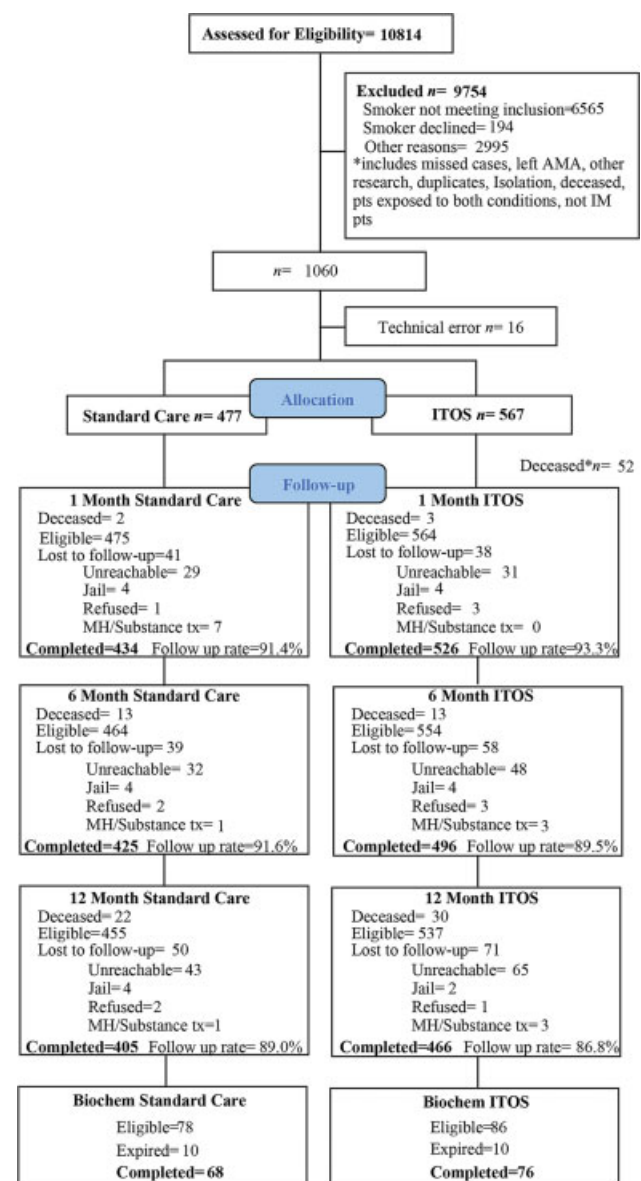- Value lists used for graphical menus.

Fig. 1 CONSORT (Consolidated Standards of Reporting Trials) diagram.

The data were then processed and integrated into a fourth table to generate the CONSORT diagram.

Enrolled participants were assigned a unique study enrollment identification (ID) and three additional IDs to correspond to follow-up time points. Baseline and follow-up data for each participant were related using the study ID as the primary key in the baseline table and the foreign key in the follow-up table. By using a secondary table to capture follow-up information for all three time points, we eliminated the need to create three times as many fields. During data export, a procedural code was run to prepare a file for data analysis that contained all participant data for both baseline and follow-up information, with the patient as the primary unit of analysis.

CONSORT diagrams require reporting how many potential participants were screened for eligibility. Our FileMaker Pro solution merged the numerical code captured in each primary screening status field (enrolled, excluded, or refused) with that in secondary screening status fields (reason excluded or time of refusal) to automatically generate total counts for each status across all screened and enrolled patients (e.g., the total number of excluded patients) (►Fig. 1). The values used to calculate counts were stored in the CONSORT table for each patient screened. During screening, if a potential participant was not enrolled, secondary status fields called *reason_excluded* and *time_refusal* were used to capture reasons for exclusion or to identify when during recruitment the individual declined participation. Other values tracked include allocation to each study arm and follow-up status at 1-, 6-, and 12-month time points.

►Fig. 2 shows the data cascade, illustrating how these values were processed to create the totals displayed in the CONSORT table. The example shows the data field calculations for a participant with a screening status of excluded (a numerical value of 1 in the excluded field). A calculated field named *total_excluded* shows the total number of patients excluded from the study for any reason, including patients who declined to participate. This field dynamically updates throughout the progression of the study.

Once an exclusion value is selected, the data entry user sees the text value (e.g., age < 18 years) while the corresponding numerical value (e.g., 0) is stored in the database record for that participant. For status fields such as *reason_excluded* or *time_refusal*, which have more than two possible value list options, additional fields corresponding to each option reflect the value in Boolean form (0 or 1).

One feature of CDMSs is the ability to use summary fields to display values such as grand totals across multiple records. For this CONSORT solution, a summary field corresponding to each Boolean field calculates a total for that option across all patient records. Again, following the example, a summary field called *total_excluded_age* returns the sum of values of the field *excluded_age* across all records. This total captures the number of screened patients excluded because they were younger than 18 years. These summary field values are stored in the screening table. Corresponding fields in the CONSORT table are then automatically set equal to the value in each total summary field.

Eligible patients who agreed to participate in the study were randomly assigned to standard care or an experimental arm, known as the Electronic Support Tool and Orders for the Prevention of Smoking (E-STOPS). CONSORT data for patients were derived following a format identical to that described previously. Examples include status at each follow-up (e.g., eligible, deceased, or lost), reasons for the participant being lost to follow-up, and total number of follow-ups completed for that time point.

The CONSORT table contains one record assigned a unique value by the application as a primary key field. This value corresponds to an equal value (foreign key) in a related field for all records in both the follow-up table and the screening table. This provides the linking mechanism by which the CONSORT table relates to the other data tables. The final stored field was an arithmetic sum of all exclusion, follow-up
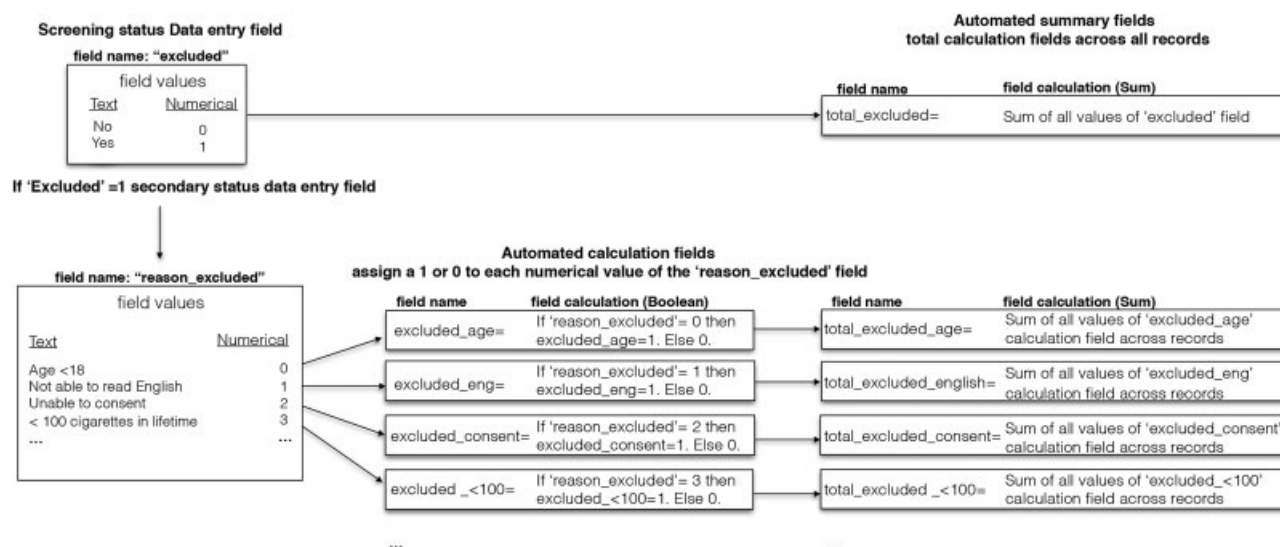
Fig. 2 Model of data entry fields and their relationship with automated CONSORT (Consolidated Standards of Reporting Trials) calculation fields.

completion, or expiration fields and was set equal to those values located in the screening and follow-up tables.

All fields used in data collection contained custom value lists stored in a value list table. Each value record had a unique text identifier as well as a secondary numerical value and other related information. This enabled user-friendly text values to be displayed in the presentation layer while storing a related numerical value in the record. These numerical values were used to automatically calculate totals for study status categories across all patient records and placed in the CONSORT table.
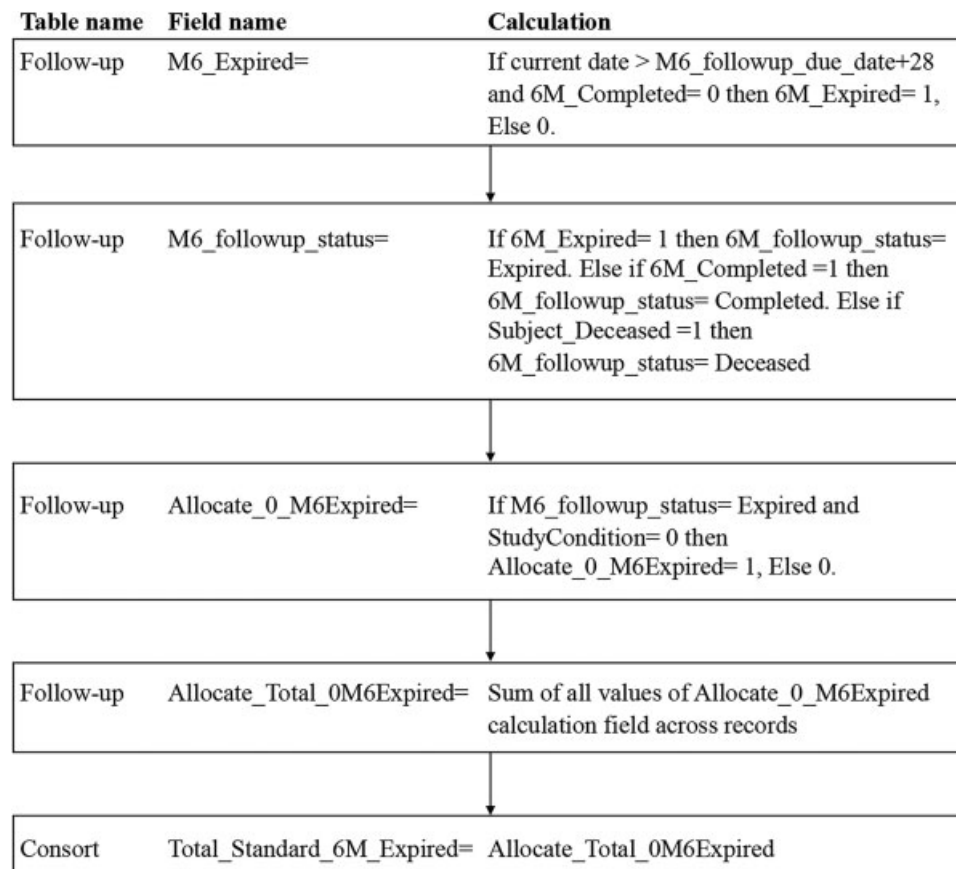
As an example, the *excluded* status field displayed text values of "yes" or "no" with corresponding numerical values 1 and 0. If a participant was excluded, then the data entry user was prompted to select a reason for exclusion. The *reasons for exclusion* value list for this study consisted of 20 text values with 20 corresponding numerical values. Only the numerical value is stored in the *reason_excluded* field of the table. This method of dynamically encoding text values into numerical values eliminates additional data entries and potential errors prior to data analysis.

Results

From August 2012 to July 2014, 10,814 patients were screened and 1,044 patients (9.7%) were enrolled. Main study results have been previously reported.[7] At 1 month, 960 (92.3%) participants were contacted by telephone; an additional 5 (0.4%) were deceased and 79 (7.6%) were lost to follow-up. Reasons for loss to follow-up were captured in a closed menu with four options: unreachable, refused to speak, incarceration, and mental health or substance abuse treatment. Once out of the time frame for follow-up, the research assistant would select one of the closed menu options. At month 6, 921 participants were contacted (90.4%); 26 participants were deceased (2.5%) and 97 participants were lost to follow-up (9.5%). At 12 months, 871 participants had been contacted (88%), 52 participants (5%), were deceased and 121 (12%) were lost to follow-up. These results were updated weekly.

Several steps are involved in calculating the data displayed in the CONSORT diagram. In some cases, up to four calculation fields are used to evaluate and store the final counts, as with participants lost to follow-up at 6 months. Over the course of the study, 39 patients in the standard care arm were lost to follow-up at 6 months (►**Fig. 1**). This total uses a function to compare the current date to the follow-up due date. Once the current date exceeds the follow-up due date by 28 days or more, the field returns the word "expired" in the 6-month follow-up status field located in the follow-up table (►**Fig. 3**). Next, a second calculation field returns a value of 1 if the follow-up record status equals expired and the study condition equals standard care; otherwise, the field returns a value of 0. In this case, 39 standard care 6-month follow-up records returned a value of 1. A third calculation

Table name	Field name	Calculation
Follow-up	M6_Expired=	If current date > M6_followup_due_date+28 and 6M_Completed= 0 then 6M_Expired= 1, Else 0.
Follow-up	M6_followup_status=	If 6M_Expired= 1 then 6M_followup_status= Expired. Else if 6M_Completed =1 then 6M_followup_status= Completed. Else if Subject_Deceased =1 then 6M_followup_status= Deceased
Follow-up	Allocate_0_M6Expired=	If M6_followup_status= Expired and StudyCondition= 0 then Allocate_0_M6Expired= 1, Else 0.
Follow-up	Allocate_Total_0M6Expired=	Sum of all values of Allocate_0_M6Expired calculation field across records
Consort	Total_Standard_6M_Expired=	Allocate_Total_0M6Expired

Fig. 3 Flow of automated calculation fields used to generate an allocated CONSORT (Consolidated Standards of Reporting Trials) diagram.

field is an arithmetic sum of the previous field across all standard care 6-month follow-up records, returning a total of 39 lost to follow-up. A fourth field in the CONSORT table is set equal to this summary field and displays the count of 39, as shown in ►Fig. 1.

Customized versions of the CONSORT diagram were created for specific device displays and printers, providing various levels of granularity depending on the study's needs. This enhanced flexibility gives investigators the opportunity to track screening and inclusion data in real time, offering insight into study operations and procedures. Importantly, this trial used an unallocated CONSORT diagram to track follow-up status during recruitment. By monitoring timely frequency counts of hospitalized individuals, important adjustments to the follow-up procedure were rapidly addressed.

Discussion

We present an approach to generating CONSORT diagrams, applicable to any prospective controlled trials. Using this approach, the study's research team was able to track the progress of participant recruitment and follow-up, understand reasons for attrition, clarify the manual of operations, and remediate research staff as needed. The tool also facilitated the gathering of data required for quarterly progress reports to the funder.

Overall, the development and implementation of CDMSs took approximately 200 hours. Development of a comparable CDMS using other relational database software could take a similar amount of time depending on what core features the research team aims to include. For this solution, the CON-SORT diagram feature required the creation of three fields for each data element. This field creation is time-intensive and took approximately 10 hours.

In a previous trial, producing a CONSORT diagram using data captured on paper case report forms and later entered into electronic spreadsheets took approximately 400 staff hours and required multiple staff members and several months to create. Multiple flat files were used because the sheer amount of data caused the spreadsheet software to continually crash. This required manual addition across files for accurate counts. In addition, data columns in flat files are not relational.[9] As a result, sorting, calculating, or manipulating any of the cells can decrease data accuracy and increase redundancy.[4]

Lastly, this trial eliminated data transcription from the EMRs by coordinating with Epic report writers to create a .csv file of all potential patients currently in the hospital. This file was imported into the CDMS using a procedural code that cleaned data for optimal data entry. This method optimized decision support during recruitment by flagging currently enrolled patients and patients already enrolled in other clinical trials and then prioritizing which patients to screen based on current length of stay as well as other inclusion factors.

Although FileMaker Pro was used to develop this solution, the summary functions (e.g., count, total) as well as Boolean logic are readily available in comparable CDMS. Given this, the creation of the CONSORT diagram described in this study could be achieved using other CDMSs.

The study has several limitations. FileMaker Pro includes several templates, but none are customized for clinical research. The figures in this study may be used as templates to generate CONSORT diagrams.

Special attention must be made to ensure compliance with federal privacy rules under the Health Insurance Portability and Accountability Act. FileMaker Pro does not include a built-in audit trail that logs all user activity. Instead, the administrator must create scripts to perform this function. Other relational databases offer an automatic de-identification feature for data export as well as an audit log.[10] In FileMaker Pro, this can be achieved through additional programming.

Lastly, our tool was developed for use in a single-site clinical trial. It can be adapted in web-based form for use in multicenter trials.

Conclusion

We offer an approach using FileMaker Pro to allow investigators to generate automatically the CONSORT diagrams that are mandatory for the reporting of prospective clinical trials. It may be adapted for use in any clinical trial. Automated generation of CONSORT diagrams is both time- and resource-efficient and allows the study team to examine the flow of patients in the trial as part of routine study management.

Clinical Relevance Statement

Investigators must report the flow of subjects participating in a clinical research trial. We provide a method to generate these figures, known as CONSORT diagrams, automatically using commercially available software. This method may save investigators time in constructing such diagrams.

Multiple Choice Question

A CONSORT diagram should include which of the following?
   a. The number of individuals screened for eligibility.
   b. A demographic profile of study subjects.
   c. A statistical test to indicate whether the intervention was effective.
   d. The number of research assistants used to enroll subjects.
   Correct Answer: The correct answer is option a. CONSORT diagrams allow the reader to track the flow of subjects in a clinical trial. This includes all individuals screened for eligibility, the number actually enrolled, the number of individuals assigned to each study arm, the number available at follow-up, and reasons for declining to participate or being lost to follow-up. The demographic profile of study subjects, efficacy of the intervention, and details about the research team are reported elsewhere in the

Methods and Results sections of this study or in other tables and figures.

Protection of Human and Animal Subjects

This study did not involve human or animal subjects. The parent trial from which this study was conducted was performed in compliance with the World Medical Association Declaration of Helsinki on Ethical Principles for Medical Research Involving Human Subjects and was reviewed by the Human Investigation Committee of Yale University.

Funding

Conflict of Interest

None declared.

References

1 Higgins JP, Altman DG, Gøtzsche PC, et al; Cochrane Bias Methods Group; Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928

2 OnCore Enterprise Research. Build a research center of excellence. Available at: https://forteresearch.com/enterprise-research-oncore/. Accessed August 18, 2017

3 REDCap. Available at: https://projectredcap.org. Accessed August 18, 2017

4 Cavenaugh JS, Snell P, Jeffries D, Waight PA, McConkey SJ. A relational database for management of flow cytometry and ELISpot clinical trial data. Cytometry B Clin Cytom 2007;72(01): 49–62

5 The CONSORT Group. Welcome to the CONSORT Website. Available at: http://www.consort-statement.org/. Accessed March 14, 2016

6 Bhatti P, Schemitsch EH, Bhandari M. Managing data in surgical trials: a guide to modern-day data management systems. J Bone Joint Surg Am 2012;94(01, Suppl 1):45–48

7 Bernstein SL, Rosner J, DeWitt M, et al. Design and implementation of decision support for tobacco dependence treatment in an inpatient electronic medical record: a randomized trial. Transl Behav Med 2017;7(02):185–195

8 FileMaker Pro. Available at: https://www.filemaker.com. Accessed July 27, 2018

9 Leff DR, Lovegrove RE, Darzi LA, Athanasiou T. Data collection, database development and quality control: guidance for clinical research studies. In: Key Topics in Surgical Research and Methodology. Berlin, Heidelberg: Springer; 2010:305–320

10 Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42(02): 377–381