

DOUBLE DATA ENTRY QUALITY CONTROL IN URECA (URBAN ENVIRONMENT AND CHILDHOOD ASTHMA), A MULTISITE LONGITUDINAL STUDY

Stephanie Hicks, Pat Zook, Rich Budrevich, Ryan Bailey, Cynthia Visness, William Taylor, Michelle Walter for the Inner-City Asthma Consortium

Rho, Chapel Hill, NC



Background/Objective

- Accuracy of data entry is essential for obtaining valid results in research studies. Double data entry (DDE) is a common quality control measure used in research studies to assess data entry error rates and improve data quality.

- The Urban Environment and Childhood Asthma (URECA) study is a multicenter prospective birth cohort study. URECA's four clinical sites use a web-based data management system (DMS) to enter all data collected during the longitudinal study.

- URECA site teams recruited mothers during their third trimester of pregnancy. Data about the mother were collected during the Recruitment and Screening periods. Once the child was born, new forms were implemented and data collection began to involve both the mother and child.

Data Collection:

- Starting in 2004, the URECA study enrolled 609 participants and the children are now between 4 and 6 years old.

- Site staff complete participant interviews and conduct study procedures. All data collected are recorded on paper copies and then entered into the individual participant's electronic forms in the DMS.

- The URECA site teams collect data during quarterly phone assessments and yearly clinic visits. The teams data enter up to 6 forms in the DMS for a typical quarterly phone assessment, and up to 34 forms for a yearly clinic visit.

Methods

Double Data Entry Process:

- The DDE process requires organization by the Statistical and Clinical Coordinating Center (SACCC) and cooperation of the clinical sites.
- The DDE exercise is performed yearly during the URECA study.



Form Selection:

- The SACCC staff identifies the forms that contain data affecting the outcome measurements of the URECA study for sampling. The accuracy of these data points is of utmost importance.
- The DMS randomly selects 5% of the outcome forms that have been entered at each clinical site since the last round of DDE.
- The DMS generates a report that lists all selected forms for each site by participant identification number, visit date and form number.

Form ID	Form Name	Form Type	User	Completed	Errors	Form ID	Form Name	Form Type	User	Completed	Errors
07-01-001-1	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-1	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0
07-01-001-2	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-2	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0
07-01-001-3	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-3	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0
07-01-001-4	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-4	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0
07-01-001-5	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-5	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0
07-01-001-6	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-6	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0
07-01-001-7	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-7	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0
07-01-001-8	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-8	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0
07-01-001-9	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-9	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0
07-01-001-10	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-10	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0
07-01-001-11	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-11	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0
07-01-001-12	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-12	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0
07-01-001-13	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-13	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0
07-01-001-14	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-14	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0
07-01-001-15	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-15	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0
07-01-001-16	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-16	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0
07-01-001-17	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-17	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0
07-01-001-18	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-18	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0
07-01-001-19	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-19	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0
07-01-001-20	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0	07-01-001-20	16-Month Clinic Visit	Form 101	Respiratory and Allergy Symptoms v1.1	10/10/09	0

Site Forms to SACCC:

- The clinical site staff locate each paper form (contained in the participant's chart) by participant identification number, visit date and form number.
- Each site team makes copies of the selected forms and sends the forms to the SACCC.

Second Data Entry:

- Two specified SACCC staff are given access to the DMS and perform second entry. As a staff member logs into the DMS, the complete list of forms selected for DDE (for all four sites) appears.
- The staff member selects the pertinent form and enters the data found on the source document.
- Once the form has been entered, the form is locked from future editing, and is removed from the list.

SACCC staff choose the participant ID for DDE forms

SACCC staff choose the form for DDE

Recruitment ID	Study ID	Site	Form Count
07-01-001-1	07-01-001-1	Site 1	1
07-01-001-2	07-01-001-2	Site 2	1
07-01-001-3	07-01-001-3	Site 3	1
07-01-001-4	07-01-001-4	Site 4	1
07-01-001-5	07-01-001-5	Site 1	1
07-01-001-6	07-01-001-6	Site 2	1
07-01-001-7	07-01-001-7	Site 3	1
07-01-001-8	07-01-001-8	Site 4	1
07-01-001-9	07-01-001-9	Site 1	1
07-01-001-10	07-01-001-10	Site 2	1
07-01-001-11	07-01-001-11	Site 3	1
07-01-001-12	07-01-001-12	Site 4	1
07-01-001-13	07-01-001-13	Site 1	1
07-01-001-14	07-01-001-14	Site 2	1
07-01-001-15	07-01-001-15	Site 3	1
07-01-001-16	07-01-001-16	Site 4	1
07-01-001-17	07-01-001-17	Site 1	1
07-01-001-18	07-01-001-18	Site 2	1
07-01-001-19	07-01-001-19	Site 3	1
07-01-001-20	07-01-001-20	Site 4	1

DMS Reports:

- The DMS generates a report showing data discrepancies between the site and SACCC entries.
- The SACCC staff compare the source document to the report to confirm that the SACCC entry is correct and re-enter data if necessary.
- Once all discrepancies have been verified, the database for that round of DDE is locked from future editing.
- The DMS generates a report presenting information about the round of DDE. The report can also be run for each individual site, including site data entry personnel. Each report shows the number of forms and fields entered, followed by the number of errors made, including:

- Overall Error Rate
- Site Error Rates
- User Error Rates
- Form Error Rates

Discussion:

- The SACCC writes a memo explaining the overall DDE exercise and results to the clinical investigators. The results are discussed on a conference call that includes the clinical investigators, clinical site staff and the SACCC. Suggestions are offered for improvement if needed.

Results

- Five rounds of DDE have been completed during the URECA study. Over the past five years, the data entry error rate has decreased.

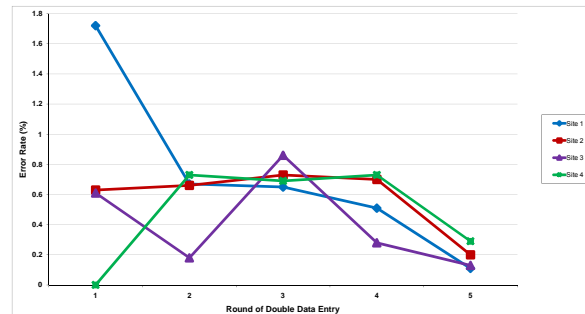
Summary of All Rounds of Double Data Entry:

	Site 1*	Site 2	Site 3*	Site 4	All Sites
Total Forms Entered	33575	27826	33063	19979	114443
Forms Sampled	1061	796	1129	618	3604
Fields Sampled	23057	18481	24898	13908	80344
Errors	165	115	119	78	477
Error Rate (%)	0.72	0.62	0.48	0.56	0.59

*Site names have been masked
*Site has dedicated data entry staff member

- A total of 3,604 forms containing 80,344 data entry fields were sampled. There were a total of 477 errors over the five rounds of DDE.

Data Entry Error Rates Over Time



- The overall error rate decreased over time.
- The first round of DDE (9,705 fields sampled) yielded an error rate of 0.95.
- The fifth round of DDE (11,649 fields sampled) yielded an error rate of 0.59.

- Site data entry rates indicate an improvement in data entry from rounds 1 to 5 with the exception of site 4.
- Site 4 had a dedicated data entry staff member at round 1 who made no errors.

Example of Outcome Form Data Entry Error Rates

	Round 1	Round 2	Round 3	Round 4	Round 5
Child Allergy Symptoms & Triggers Update*			2.47	0.18	0.00
Respiratory and Allergy Symptoms	4.65	0.68	0.34	0.43	0.00
Respiratory Illness Scorecard	0.72	2.33	2.43	0.61	0.25

*Form was not used until Round 3

- The error rate for the outcome forms helped to identify complicated forms.
- The SACCC discussed the form problems with the site teams and modified the forms when possible.
- The error rates for these identified forms decreased over time.

Example of Data Entry Error Rates for Individual Staff Members

	Round 1	Round 2	Round 3	Round 4	Round 5
Staff A	1.54	2.61	0.00	0.00	
Staff B	1.49	3.19	2.91	0.00	
Staff C	0.18	0.13	0.53	0.16	0.00
Staff D		0.00	0.67	1.48	0.86

- The overall error rate for many staff members has decreased as the study progressed.
- The error rate decreased after round 2 for most staff members.
- Staff member D had a higher error rate than the other staff members. D is a lab tech entering only lab forms, which are typically more difficult than other forms.

Discussion

- The first URECA child was born in February 2005 and the first round of DDE occurred in January 2006.
- The first round of DDE captured forms primarily pertaining to maternal data.
- Additional forms were implemented to capture data about the child after the child's first birthday, and each successive year.
- The increased error rate seen from round 1 to round 2 is likely due to the new forms.
- DDE identifies many of the difficult forms and reveals the staff members who may have trouble with data entry. Detecting these problems allows for the SACCC and the sites to implement changes to ensure quality data is entered and analyzed.
- The DDE exercise in URECA is different from many other studies, as the double entry is performed by the SACCC. Performing the exercise in this way decreases the burden on the site teams.
- There is still work involved for the site teams, as they must locate the forms in the participant binders, photocopy each form, and send the forms to the SACCC, while maintaining daily study responsibilities.
- The SACCC gives the site teams one month to complete this process.

Conclusion

- The URECA study will continue to use DDE as a quality control measure. Although all errors will not be detected through this measure, DDE allows for errors to be detected in a sample of outcome data.
- The low overall error rate suggests that much of the URECA data is clean and accurate, and the comparison of error rates over time shows that sites continue to improve in data entry. The improved error rate could be due to increased staff familiarity with the forms over time or revisions to complicated forms to improve staff usability. The DDE process may also play a role in error rate improvement. Staff may perform better because of the feedback provided during DDE discussions or a tacit desire to reduce their error rate from round to round of DDE.
- Some researchers¹ have found DDE to be valuable in producing quality data while others have not.^{2,3} S. Goldberg, et al.³ conclude that further investigation is needed for the detection and prevention of data entry errors. In the URECA study, we have found that double data entry is one of many techniques for ensuring quality outcome data. When implemented successfully, it can improve data integrity, help identify difficult forms, and reveal struggling staff who need more data entry training.

¹Remington-Harris RA, McBride R. Single vs. Double data entry in CAIST. Control Clin Trials 1992; 13:487-494.
²Glaser D, Harris AJ, Engel B, Thomas MK. Is double data entry necessary? The Crystal Field. Control Clin Trials 1996; 16:400-408.
³Goldberg S, Nimmervoll A, Turton A. Analysis of Data Errors in Clinical Research Databases. AAAA Annu Symp 2008; 242-248.

