

Society for Clinical Trials 2016

Characterization of Double-Data-Entry in Data Entry Systems Employing Modern Interface Elements

Kristen Kaiser
Center for Clinical Trials
Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health

18-May-2016
Montréal, Quebec

Data System

- JHCCT web-based data management system
 - 2001-present
 - ~3 dozen multi-center trials (and other studies)
 - Supports networks (multiple simultaneous trials)
 - Permits, but does not require, double-data-entry, at the option of the study leadership
 - Makes heavy use of “modern” interface elements

Data System

Text, Numeric, and Date Items

3. Participant code:

4. Date of interview:

Data System

Multiple-choice (choose all that apply), implemented as associated checkboxes

29. What was amputated
(check all that apply)

- (a) Toe
- (b) Leg
- (c) Arm
- (d) Other *(specify)*

Data System

Single-choice (choose only one) item, implemented as radio buttons (and showing skip pattern)

15. Overall, how has your weight changed over the past year:
(check only one)
- (1) Stayed about the same → 17.
 - (2) Gained weight → 17.
 - (3) Lost weight
 - (4) Both gained and lost weight

Data System

Single-choice (choose only one) item, implemented as a drop-down box



1. Center ID:

ZTST

<no selection>

GWCC

PROH

LAB

KAI

CO

DCC

DSMB

ZTST

The image shows a screenshot of a data entry form. The form has a yellow header bar with the text '1. Center ID:'. Below the header is a large white text area. To the right of the text area is a drop-down menu. The menu is currently open, showing a list of options: '<no selection>', 'GWCC', 'PROH', 'LAB', 'KAI', 'CO', 'DCC', 'DSMB', and 'ZTST'. The 'ZTST' option at the bottom of the list is highlighted with a blue background. A small downward-pointing arrow is visible in the top right corner of the menu box.

Double-data-entry

- Double Data Entry (DDE)
 - Elements of the Interface
 - Immediate Dependent
Pass 1 and Pass 2
- Types of Error
 - True – Keep Pass 2 Entry
 - False – Keep Pass 1 Entry

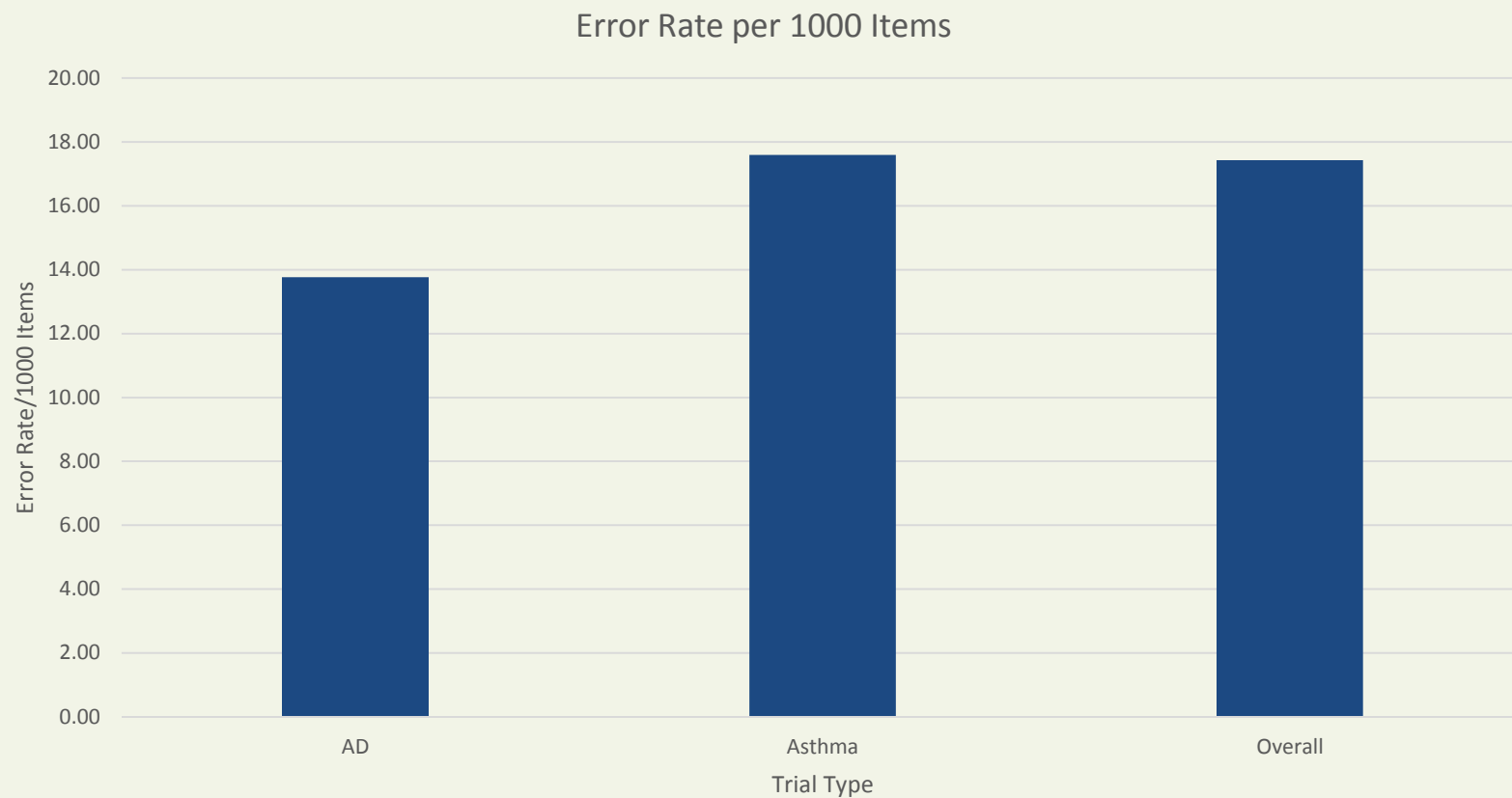
Studies

- 7 Complete Trials
 - 2004-2015
 - 5 Asthma, 2 Alzheimer's Disease (AD)
 - Multi-Centre
 - 60-600 Participants
 - Range of Age Groups
 - Ignored Subsequent Edits
- 174 Case Report Forms (Average of 52 Items/Form)
- 226,894 Total Records
- 14,242,577 Total Items Entered

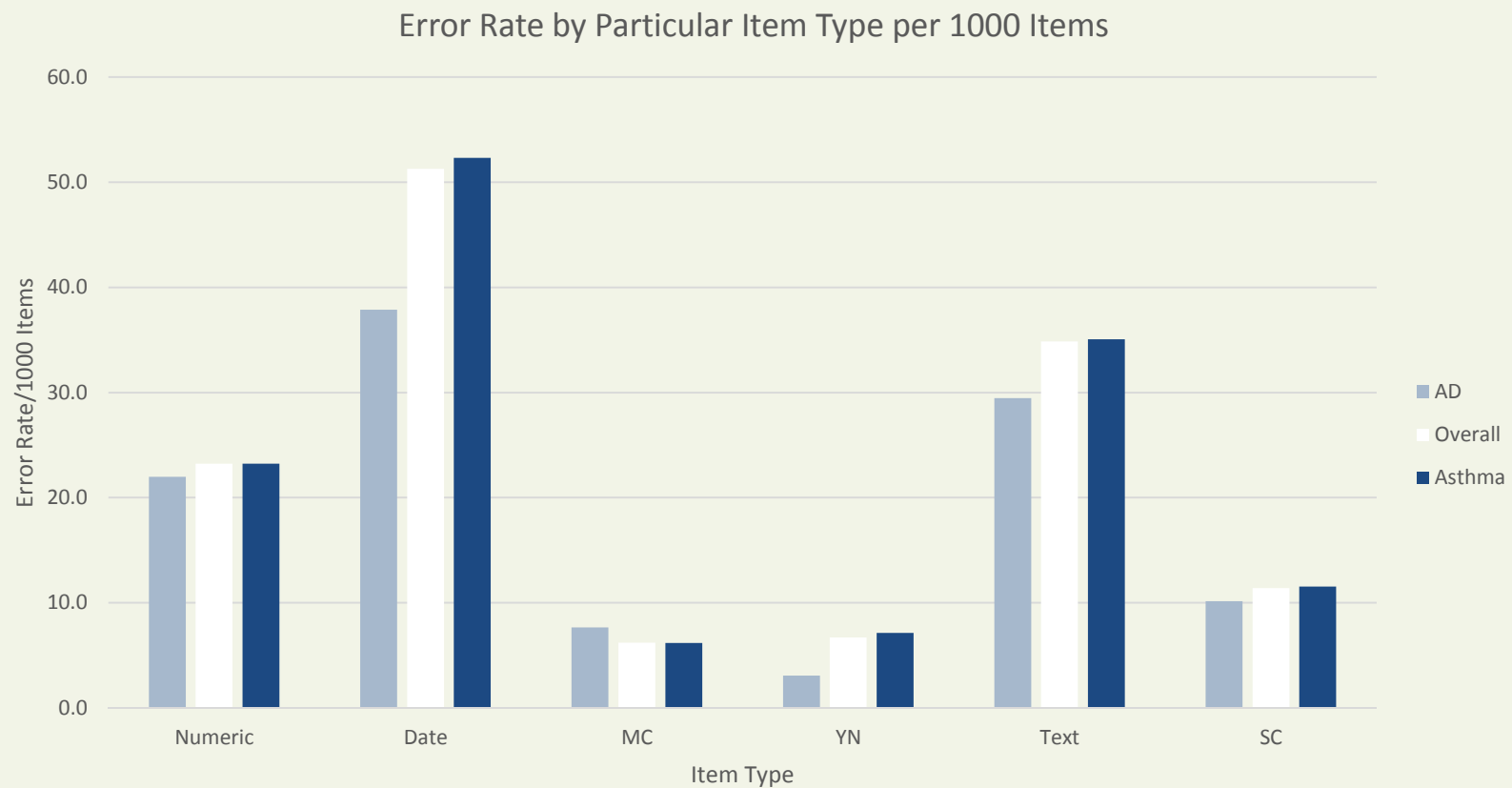
DDE Errors

- 248,224 Total Errors
 - 45,720 “True” Errors (18.4%)
 - 202,504 “False” Errors (81.6%)
- 17.4 Overall Error Rate per 1000 items
 - 13.2 False Error Rate per 1000 items
 - 3.2 True Error Rate per 1000 items
- 42.9% of Total Records contained some DDE Error
 - 68% of records that had at least one error contained only false errors

DDE Error Rate



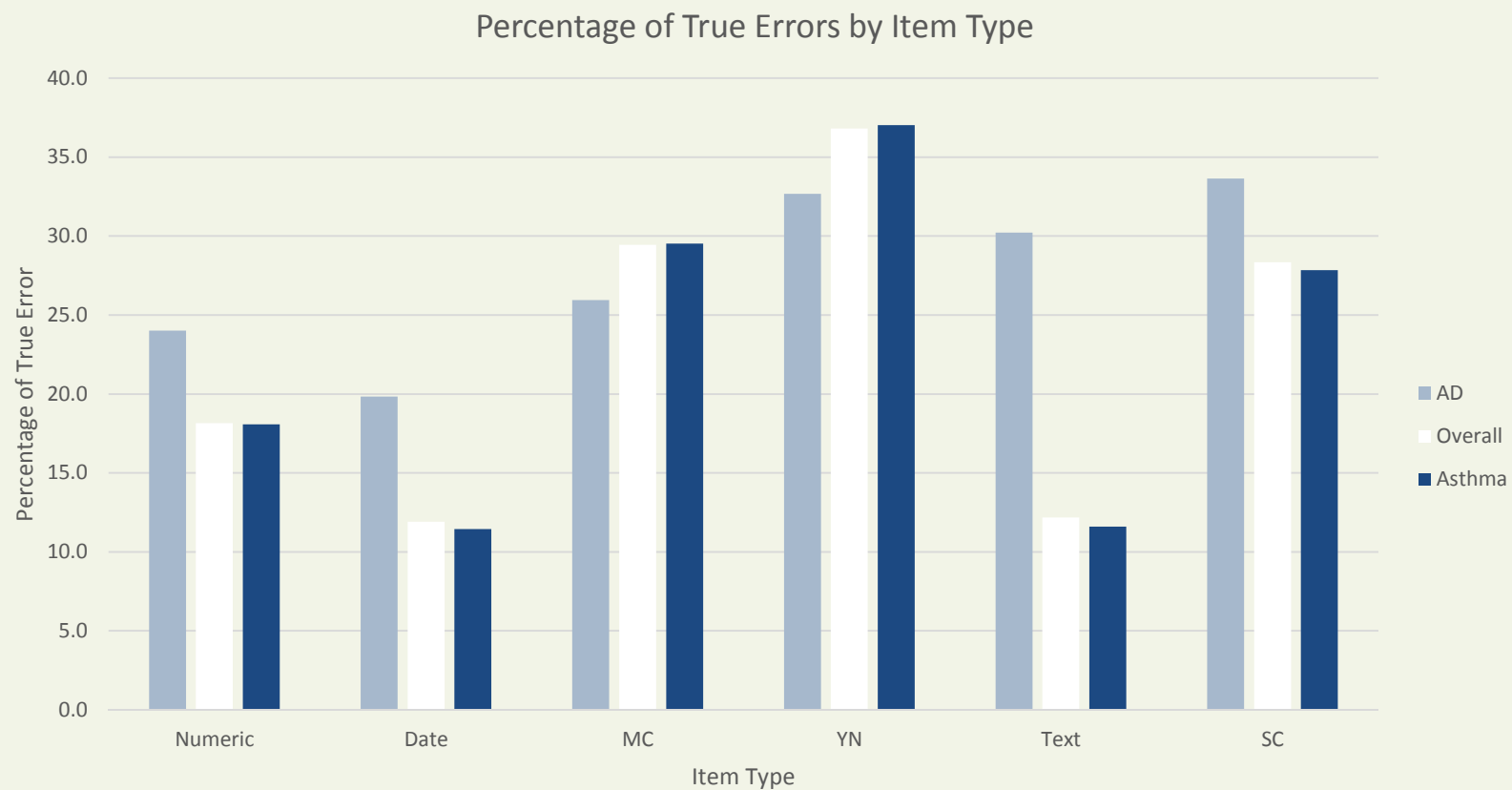
DDE Error Rate



DDE True Errors



DDE True Errors



Discussion

- Overall error rates roughly consistent with previous work
- Vast majority of DDE errors were not helpful in improving data quality
 - DDE introduces over 80% of errors – “false errors”
- Directly keyed item types triggered DDE errors at much higher rates
 - Radio-buttons showed lower rates of errors but higher percentages of True Errors

Discussion (cont'd)

- Our DDE system only captures entry or keying errors
 - May capture some transcription errors – mistakes in reading data written on CRFs
- Most forms are entered without error
 - Some forms – mostly complex – much more prone to DDE errors

Conclusion

- DDE may not be the most efficient way to detect keying errors
 - 45K “true” errors: is this worth 14M DDE keyings?
- Are there other ways to assure entry quality?
 - Making simpler forms – and perhaps splitting complex forms – may improve quality

Future

- Relationship between form length (# of items) and data entry quality
- Analysis of subsequently hand-audited results to investigate the reduction in entry errors achieved by DDE

Questions?

- Thanks to Dave Shade and Andie Lears for their help with this project and for the opportunity to be here.