Perceived Reasons for High and Low Quality Observational HIV Research Data

Stephany N. Duda\textsuperscript{a}, Catherine C. McGowan\textsuperscript{a}, Cynthia S. Gadd\textsuperscript{b}

\textsuperscript{a}Department of Biomedical Informatics, Vanderbilt University School of Medicine, Nashville, TN, USA
\textsuperscript{b}Department of Medicine, Division of Infectious Diseases, Vanderbilt University School of Medicine, Nashville, TN, USA

Abstract

Audits of data quality in a Latin America HIV research network revealed that study sites collected weight measurements, laboratory results, and medication data of inconsistent quality. We surveyed site personnel about perceived drivers of their high or low quality data. Most sites reported their research teams contained no data specialists and that missing data stemmed primarily from incomplete patient assessments at the point of care rather than inconsistent data recording. The root causes of data errors resulted from limited clinic resources (e.g., broken scales, limited record storage space), workflow complications, or the indifference of external participants towards research activities. Understanding these factors supports targeted quality improvement processes.

Keywords:
Data Collection, Error Sources, Data Quality

Introduction

The Caribbean, Central, and South America Network for HIV research (CCASAnet), a network of HIV clinics in Latin America, conducts on-site reviews of routine clinical care data submitted for research by study sites. During a data audit, data coordinating center personnel check data submitted for research against the contents of paper clinical records (source documents), flagging data mismatches, missing data, and data with no documented origin. This source document verification helps quantify data quality to ensure meaningful research findings [1]. The audits revealed many discrepancies in CCASAnet sites’ data. We investigated the causes of data quality differences across sites through a data quality survey.

Materials and Methods

We developed a survey that asked site personnel about the presumed causes of their high or low quality data. We focused on 3 high-error categories: weights, lab values, and medication history. The survey was approved by the Vanderbilt IRB, implemented in REDCap (projectredcap.org), and distributed by coordinators at each site. Survey responses were grouped by site, but anonymous within sites. The project ran March-June 2009. Two researchers independently selected recurrent themes through iterative, bottom-up analysis. Conclusions were harmonized through group discussion. A third researcher was appointed to resolve any conflicting interpretations.

Results

We received 18 survey responses, representing 6/7 CCASAnet sites. Respondents reported that clinicians conducted data abstraction rather than data personnel, and teams included on average 3-5 people. Reported reasons for missing weight data included broken scales and providers’ perception that weight measurements were unimportant for clinical care. Two sites with high quality weight data described reducing missingness via multiple data capture opportunities. Auditors often could not find source documentation for the laboratory data sent for research. Survey responses described oversized laboratory printouts that didn’t fit into medical record folders, a lack of shelf space in the record room, and the practice of giving lab reports to patients as reasons for missing paper documents. The one site with high quality lab data received direct data exports from the laboratory system. Sites with low quality medication data agreed that brief drug regimens were easily overlooked during data abstraction and delays between when a physician prescribes treatment, the government approves the treatment, and the pharmacy dispenses the drugs resulted in multiple, conflicting dates recorded in the patient chart. Collaborative work with local clinic pharmacies was reported as a reason for higher quality data, and sites benefitted from having access to pharmacy records as a secondary source of information.

Discussion

The responses of survey participants revealed that higher quality research data was the result of complementary processes (e.g., parallel laboratory or pharmacy workflow) or automated data imports, which reflects well-documented quality improvements when reducing manual transcription [2]. Most on-site data preparation roles were filled by clinicians although trained data clerks may conduct accurate and more cost-effective abstraction of basic data [3]. Respondent feedback informed potential approaches to reduce errors in HIV research data abstracted from patient care records.

Acknowledgments

The authors thank Daniel R. Masys, MD, and the teams at the CCASAnet sites. This work was supported by NIH/NIAID U01AI069923 and NCATS UL1TR000445.

References


Address for correspondence
Stephany Duda, stephany.duda@vanderbilt.edu