



N = 1: The inaccuracy of a non-validated clinical data registry.

Wendi R. Mason, MSN, RN, ACNP-BC, James Del Greco, MSN, RN, Eric Smith, LPN, and Susan Martin, LPN
Vanderbilt University Medical Center, Department of Medicine, Division of Allergy, Pulmonary, and Critical Care

Introduction

Data integrity refers to maintaining and assuring the accuracy and consistency of data. Data validation is the pre-requisite for data integrity and involves the cleaning of data to ensure high quality and accurate information. Clinical data registries are tools that allow the capture and analysis of data and usually focus on a specific population, procedure, or treatment, includes snapshots or longitudinal parameters, to understand various outcomes. Often practice patterns are modified based on findings of clinical registries.

Idiopathic pulmonary fibrosis (IPF) is a chronic restrictive fibrotic lung disease of unknown etiology and is associated with oxygen desaturations, hypoxemia progressing to respiratory failure and death. Nearly every academic medical center providing healthcare to IPF patients has built and maintains a patient registry and there are a growing number of national IPF and interstitial lung disease (ILD) patient registries sponsored by pharmaceutical companies, non-for-profits, and others, most of which do not provide a method of external data validation.

Purposed / Hypothesis

The purpose of this study was to determine how correct a unique patient would be characterized when clinical data was entered into a unmonitored and non-validated data registry.

The hypothesis was that non-validated unmonitored data would be an inaccurate representation of the patient snapshot.

Methods



Four research study nurses experienced in the diagnosis were asked to fill out a 2-page case report form (CRF). Each nurse was given brief instructions about the protocol and the CRF, completed the CRF by retrospective chart review, were not allowed to ask for help, and were instructed to time themselves.

The paper CRF contained 13 sections (71 entry items) characterizing the demographics, medical history, exposures, and physiologic status of a patient.

The patient was randomly selected and met the inclusion and exclusion criteria of a proposed data registry. He is an established patient in the program and well known to the four research nurses. The visit to be recorded had occurred four months prior.

The research director scored the CRFs for correctness, incorrectness, and omissions of data. Statistics were descriptive.

Results

Based on the retrospective review, the unique patient was recorded in the data registry as having four very different characteristics related to IPF.

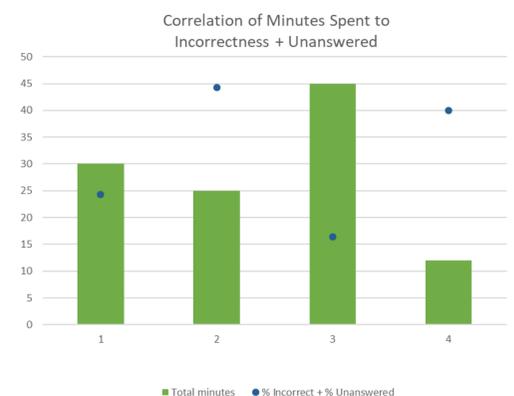
Of the 13 sections of the CRF, only two sections (family history and pulmonary function testing) were correctly recorded by each nurse.

Sections of the CRF that were inconsistently reported by 3 of the 4 nurses included demographics (height and body mass index), use of pulmonary rehabilitation, smoking history (ever smoker), and supplemental oxygen liter flows. Omissions of data occurred by 3 of the 4 nurses in the sections recording smoking history (year quit, packs per day) and the six-minute walk. 50% of the laboratory data was incorrectly reported or omitted.

Excerpt of Score Card showing Percent Correct

	Nurse 1	Nurse 2	Nurse 3	Nurse 4	# correct	% correct
Oxygen	1	2	1	1	3	75
cont	2	0	1	0	1	25
sleep	2	0	1	0	1	25
exert	2	0	1	2	1	25
exert and sleep	2	0	1	0	1	25
rest LPM	2	0	1	0	1	25
sleep LPM	2	0	1	0	1	25
Exert LPM	1	0	1	1	3	75
Ever smoker	1	2	2	2	1	25
Quit yr	0	0	0	0	0	0
ppd	0	0	0	0	0	0
# yrs	0	0	0	0	0	0
Current smoker	1	1	1	0	3	75
6MWT Data date	2	2	1	2	1	25
oxygen	0	0	1	0	1	25
liter	0	0	0	0	0	0
Pre borg	0	0	1	0	1	25
Post borg	0	0	1	0	1	25
distance	0	0	1	0	1	25
# rest	0	0	1	0	1	25
hr at end	0	0	1	0	1	25
hr at 1 min	0	0	0	0	0	0
lowest sat	0	0	1	0	1	25

0=no answer, 1=correct answer, 2=incorrect answer



Nurse 3 was most correct (77%), least incorrect (14%), and had the least data point omissions (6), but spent the most time completing the CRF (45 minutes).



Nurse 4 spent considerably less time completing the CRF and scored the highest incorrectness + omissions percentage.



Discussion

All registries can provide useful information, but there are internal rigorous requirements that should be mandated in order to assure valid and accurate data regardless of the purpose of the registry.

Several issues were identified as reasons for inconsistency and omissions in the data collection.

1. Electronic medical record (EMR)
 - Multiple locations of the same data were inconsistently recorded throughout the medical record
 - Internal limitations blocked access to data for some users
2. CRF – The poor design of the CRF did not allow for intuitive data capture
3. Knowledge base – lack of understanding of the objective of the study

An internal (2nd reviewer) audit of this data set is considered the best tool to ensure a high level of quality and accuracy of the submitted data. Internal monitoring provided by the sponsor would only validate what the completer entered. Database ranges set for outliers would not have captured the inconsistencies.

Limitations

- Nurse 1 participated in the study as one of the completers of the CRF, but also randomly selected the patient and was also the one who scored the four CRFs. While Nurse 1 did not score herself the most correct, the supposition exists that she believes her answers to be the most correct, whereas if a non-completer scored the CRFs, the correct and incorrect scores may have been different.



Conclusions

Clinical data registries do not always fall under the strict guidelines of data validation, as are set forth in randomized clinical trials. Data registries are important tools that through analysis and statistics guide researchers and clinicians in understanding patient outcomes, and precipitate change in how healthcare and treatments are provided to patients. In order to provide the most accurate data, assurances for data validation and monitoring should be part of the initial design of every clinical data registry protocol.

Acknowledgements

I would like to thank the four research nurses who allowed complete transparency in the presentation of this data.