Attachment 1 to: Stausberg J, Bauer U, Nasseh D, Pritzkuleit R, Schmidt CO, Schrader T, Nonnemacher M. Indicators of data quality: Review and requirements from the perspective of networked medical research. GMS Med Inform Biom Epidemiol. 2019;15(1):Doc05. DOI: 10.3205/mibe000199

Definition of indicator completeness[1]

Description

Completeness of the data

Definition of terms

Completeness: Extent to which data that could be recorded had been really recorded [Arts et al. 2002a].

Identifier

TMF-1046

Type of indicator

Outcome

Literature references

Arts et al. 2002a, Barrie/Marsh 1992, Bobrowski et al. 1999, Goldberg et al. 1980, Hassey et al. 2001, Hogan/Wagner 1997, Jensen et al. 2002, Jung/Winter 2000, Katalinic 2005, Kuntoro et al. 1994, Lindquist 2004, Logan et al. 2001, Naumann/Rolker 2000, Nielsen et al. 1996, Parkin/Muir 1992, Teppo et al. 1994, Topp et al. 1997, Vestberg et al. 1997

Context

Individual record, individual observational unit, complete data set. Calculation for data elements.

Alternative definitions

Completeness could be alternatively defined as 1) percentage of patients/subjects with complete data from all patients/subjects or 2) sensitivity [Hassey et al. 2001, Nielsen et al. 1996], i.e. percentage of patients/subjects that exhibit a recorded characteristic from all patients/subjects that exhibit this characteristic.

Comments

This indicator is important for a correct estimate of incidence and prevalence rates. The indicator is related to the indicators concordance and “data elements with unspecific values”. Completeness of cases is captured with the indicator “recruitment rate”. Reference for an estimate of completeness could be sources as other registries, death certificates, administrative data, and others.

Numerator

Number of available values [Naumann/Rolker 2000]

Denominator

Number of available values + Number of missing values (= Number of values, that could be recorded) [Naumann/Rolker 2000]

Subcategories

None

Method of calculation

1) Identification of data that should be recorded

2) Check, whether the data are available

3) Calculation of numerator, denominator, and rate

[1] The literature references of the example are not included in the paper.

Interpretation of results

The higher the rate, the better the data quality. The threshold depends on the intended data use.

Predictors and confounders

Available time and resources for data recording

qualification of the staff responsible for data recording

completeness of documents that are used as source for data recording