

Table 1: Categories of potentially identifying variables reported in the sample of articles

Identifier	Description
HIPAA safe harbor	
HIPAA safe harbor	Any of the 18 HIPAA safe harbor identifiers: names; geographic subdivisions smaller than a state; all elements of dates (except years) for dates that are directly related to an individual (including birth date, admission date, discharge date, death date, and all ages over 89); telephone numbers; vehicle identifiers and serial numbers, including license plate numbers; fax numbers; device identifiers and serial numbers; e-mail addresses; Web Universal Resources Locators (URLs); social security numbers; internet protocol (IP) addresses; medical record numbers; biometric identifiers, including finger and voice prints; health plan beneficiary numbers; full-face photographs and any comparable images; account numbers; certificate/license numbers
Non-HIPAA safe harbor	
Geographic location	Any reference to a geographic area that is not categorized as a HIPAA safe harbor identifier; includes any reference to a country, a US state, or an international equivalent of a state or similar-size region
Institution name/organization	Any reference to the name of an institution or organization
Race, ethnicity, indigenous status, or nationality	Any reference to race or ethnicity, as well as indigenous groups and tribes, and country when used to indicate the nationality of an individual
Age	Any reference to the age or age range of a person that is not categorized as a HIPAA safe harbor direct identifier of age (includes references to age in years, months or weeks but not birth dates which would be considered a HIPAA direct identifier)
*Profession or professional title ¹	Any professions and jobs including titles and qualifications
Rare disease or illness	Any reference to a rare disease that is contained in https://globalgenes.org/rarelist/
LGBTQI/sexual orientation	Any reference to sexual orientation or gender identity that is not heterosexual or cis
*Foreign language or dialect	Any reference to a foreign language or dialect as it pertains to individual participants
*Sex or gender	Includes only explicit references to sex or gender
*Education level	General reference to education level attained by participant (elementary, trade school, etc.)

¹ "****": These identifiers are not included in our current software.

Non-HIPAA safe harbor	
*Marital status	General reference to marital status of participant (single, widowed, etc.)
*Type of study site: single, multi-site, or generalized/broad community	<p>Single: research took place at one particular institution or organization</p> <p>Multi-site: research took place at two or more particular institutions or organizations</p> <p>Generalized/broad community: research took place within a generalized or broad community setting (participants with a particular disease, nursing students in Australia, etc.)</p>