116TH CONGRESS
2D SESSION

S. 4990

To require the Office for Civil Rights of the Department of Health and Human Services to conduct a study and issue a report on the de-identification of data pursuant to privacy regulations.

IN THE SENATE OF THE UNITED STATES

DECEMBER 9, 2020

Mr. CASSIDY introduced the following bill; which was read twice and referred to the Committee on Health, Education, Labor, and Pensions

---

A BILL

To require the Office for Civil Rights of the Department of Health and Human Services to conduct a study and issue a report on the de-identification of data pursuant to privacy regulations.

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE.

This Act may be cited as the ''Health Data De-identification Report Act''.

SEC. 2. STUDY AND REPORT BY THE HHS OFFICE FOR CIVIL RIGHTS.

(a) IN GENERAL.—The Office for Civil Rights of the Department of Health and Human Services, in cooperation with the National Institute of Standards and Technology, shall conduct a study that considers and compares the effectiveness and validity of safe harbor and expert determination as methods of carrying out the de-identification of data required under section 164.514 of title 45, Code of Federal Regulations (as in effect on the date of enactment of this Act), and shall submit to Congress a report on such study not later than 270 days of the date of enactment of this Act.

(b) CONTENT OF STUDY AND REPORT.—The study and report under subsection (a) shall consider—

(1) any known instances where data that was de-identified as required under section 164.514 of title 45, Code of Federal Regulations (as in effect on the date of enactment of this Act) became re-identified, even partially, and, with respect to each such instance, how such re-identification occurred and any harm incurred by any individual whose data was re-identified;

(2) the frequency by which each of the safe harbor method and the expert determination method is used for the de-identification of data as described in

subsection (a), and the sizes of the data sets on which each such method is used;

(3) the costs of each such method of de-identification, and the relative utility of the data that is de-identified through each such method;

(4) which of the 2 methods of de-identification renders data less likely to be easily re-identifiable;

(5) whether either such method removes information that would help address health care disparities, thereby making it more difficult to address such disparities;

(6) which of the 2 methods is most commonly used for purposes of medical research, and the benefits and disadvantages of using each such method for such purpose;

(7) the risk of re-identification by each of the 2 methods, especially when the resulting de-identified dataset will be accessible by entities that have access to additional data;

(8) whether or not there are use cases that cannot be performed using data de-identified by either of the 2 methods; and

(9) the ease or difficulty of expert determination, taking into consideration cost, accessibility, and qualifications of experts.

○