**Supplement**

The protocol for linking cancer diagnosis data from cancer registries to pregnancy outcomes in vital records was developed by the Georgia Cancer Registry and applied uniformly across all three states. The goal was to develop a linking strategy that limited the number of possible matches requiring manual review to a manageable number, while still allowing women’s birth and cancer records to link with slight variations such as different spellings of the same name or transposition of digits in the Social Security number. The Georgia Cancer Registry developed a two-step method that first used a deterministic link to narrow the population to likely matches, then a probabilistic link that scored pairs of records on their likelihood of being a true match.

Cancer registry staff in each state used Link Plus, a record linkage program developed by the U.S. Centers for Disease Control and Prevention, to conduct the link. Registries were only able to link pregnancies to cancer records within their state. Thus, we could correctly identify a cancer survivor who delivered in Georgia and was diagnosed in Georgia, but not a woman who delivered in Georgia and was diagnosed in Tennessee.

Before the match, live birth and stillbirth files were concatenated into one file, which allowed us to identify every pregnancy that reached 20 weeks, whether or not it ended in live birth. For the initial deterministic step of the link, cancer registries assigned an identification number (ID) to each woman that took the format: first two letters of first name, first two letters of surname, last two letters of surname, and 8-digit date of birth. For example, the ID for a fictional woman named Ada Lovelace, born October 23, 1980, would be adloce10231980. Because we wanted to capture women who changed their surnames between the time of cancer diagnosis and birth, staff created two separate IDs for women who had more than one surname listed in vital records. For example, a woman with a married name of Ada Lovelace and family name of Ada Smith would be assigned two IDs: adloce10231980 and adsmth10231980.

In the first, deterministic step, we required women’s vital record and cancer record to match exactly on either ID or Social Security number. Women whose records were an exact match on either of these two variables were retained for the second step, which was a probabilistic link. In the second step, the population was limited to women who matched in the first deterministic step. These women were then linked probabilistically on their date of birth, Social Security number, first name, middle name, and all last names (family name in vital records to family name in cancer record, and married name in vital record to family name in cancer record).

Link Plus scores the probability that each linked pair is a true match. We accepted all pairs scored ≥16 as true matches and rejected those with scores <7. For matches with scores between 7 and 16, cancer registry staff did a manual review considering census tract, zip code, race, ethnicity, and for deceased women, date of last contact. For deceased women, if the date of last contact was before the before the birth or fetal death, the match was determined to be false.

Identifying data including names, Social Security numbers and geographic identifiers were removed before cancer registries released the data for analysis.