Pwned The Risk of Exposure From Data Breaches∗

The study field:

Research on the frequency of Americans' private information being part of data breaches is limited. Most studies focus on the impact on corporations and people, but there is a lack of data on the average American's risk of their private information being exposed. This study uses a unique dataset to estimate the lower bound of the average number of breached online accounts per person. The researchers merge data from a large representative sample from YouGov with data from Have I Been Pwned, checking if the email associated with the YouGov account is part of the 293 public breaches cataloged by HIBP. The study also examines how exposure to breaches varies by socio-economic factors.

Result:

The number of breaches associated with the 5,000 emails on file is 14,979, with an average of three breaches per person. At least 82.84% of Americans' accounts have been breached at least once. The relationship between the number of breaches and socioeconomic status is contradictory to traditional concerns about the digital divide. The data suggests that people who use online services more are somewhat more likely to have their accounts breached. The number of breaches increases monotonically with education, with the average number of breaches among people without a high school degree being 2.35. In contrast, the relationship between the number of breaches and age is curvilinear, with young people's and seniors' accounts least likely to be breached and middle-aged adults' accounts most likely.

The study reveals that women's accounts are 1.12 times more likely to be breached than men's, and that Blacks and Whites' accounts are most frequently breached. The mean number of breaches associated with emails for Blacks and Whites is 3.12 and 3.16 respectively, while Hispanics/Latinos have 2.5 and Asians have 2.82. The 14,979 breaches stemmed from 156 sites, with 21 sites having more than 100 breaches accounting for 11,783 of the breaches. The analysis does not distinguish between different types of breaches, but not all are equally grave.

The study reveals that out of 15,837 breaches, 94.58% were verified, with about a third being classified as SpamList. The study focuses on the smaller subset of breaches, examining the relationship with socio-economic variables. Education shows a higher number of breaches per person for those with a college degree or more. The pattern is more attenuated for women, with the mean for women and men being 2.15 and 2.05 respectively. The general pattern for age remains similar, with middle-aged individuals more likely to have their accounts breached compared to those younger than 25 and older than 65. Breaking down by race, Asians join Whites at the top of the pile, with means of about 2.2. Hispanics or Latinos are less likely to be part of verified non-spam-list breaches, while African Americans are more likely to be part of unverified, SpamList breaches.

[Paper link](https://dl.acm.org/doi/10.1145/3292522.3326046)