

“
The propensity for misreporting and skewing datasets increases with sensitive topics such as sexual and mental health, or topics with social stigma such as smoking, alcohol and drug use.
 ”

Economical with the truth

By Kitty Drok

When surveyed about sensitive topics such as sexual health or drug use, it's understandable that some participants may provide unreliable answers, skewing survey results. Professor Mark Harris is using econometrics to address such misreporting.

Some datasets are objective, like recorded clinical incidences of diabetes. Others are compiled from self-assessed surveys, and the data is inherently subjective. Misreporting is well known for such self-reported data: when comparing the difference between measured and self-reported height and weight, women tend to under-report their weight; and younger men over-report their height more than women the same age, for example.

More significantly, around 75 per cent of the Australian population self-assess their health as 'good' or 'very good'. But clinically more objective measures of health find that 60 per cent of Australians are mildly overweight or obese, 50 per cent have elevated cholesterol, and almost 25 per cent have either diabetes or impaired glucose metabolism.

The propensity for misreporting and skewing datasets increases with sensitive topics such as sexual and mental health, or topics with social stigma such as smoking, alcohol and drug use.

Professor Mark Harris from Curtin Business School is using econometrics, the application of probabilistic models to describe real-world trends, to address misreporting in large health surveys.

When modelling social 'bads' such as illegal drug consumption, researchers often see an excessive amount of 'zero' (no consumption) observations. Harris uses statistical methods to assess survey participants' individual reporting behaviour, gleaned from their answers to (superficially) unrelated, non-sensitive questions. He uses this information to calculate the likelihood that their zero answer genuinely represents non-participation, deliberate misrepresentation (lying about participation), or infrequent participation (those who have participated at other times or in other circumstances).

"You can calculate the different probabilities all leading to the same zero result, based on other survey variables," explains Harris. "We can't validate our analysis against hard data, but the confidence intervals around our results are very tight. We work with very large survey sets to ensure the results are statistically relevant."

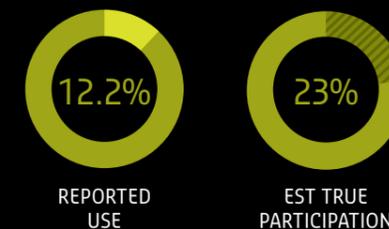
In the case of illegal drug use in Australia, the reported participation rates of 12.2, 3.2 and 1.3 per cent for marijuana, speed and cocaine respectively, mask true participation rates estimated to be almost double for marijuana (23 per cent), and more than double for speed (eight per cent) and cocaine (five per cent).

Harris is blunt about the bottom line: "It's not sensible to look at these raw, self-reported aggregate numbers and assume they accurately represent what's going on in the population. Even small differences in corrected percentages can translate into large numbers of people. Systematic misreporting has significant implications when this data is used as a basis for policymaking or health funding decisions."

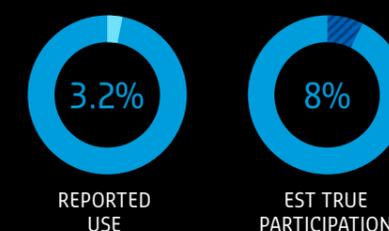
business.curtin.edu.au

Illegal drug use: the truth

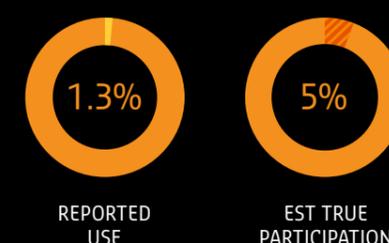
Marijuana



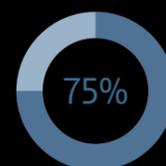
Speed



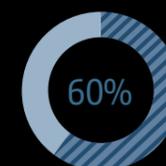
Cocaine



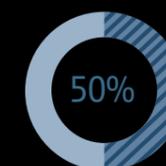
Overstating our health



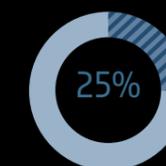
SELF-ASSESS THEIR HEALTH AS GOOD OR VERY GOOD



ARE MILDLY OVERWEIGHT OR OBESE



HAVE ELEVATED CHOLESTEROL



HAVE DIABETES OR IMPAIRED GLUCOSE METABOLISM