Which lottery numbers do players play?

The Irish national lottery went through 62 rollovers during 2021 and 2022 before a winner finally claimed the top prize. **James Hanley** and **Michael Cronin** use data from four continents to examine what lies behind lottery droughts, and estimate the most, and least, popular numbers being picked

The 2021 jackpot drought in the Irish Lottery received considerable comment from the media, statisticians, Irish politicians and lottery administrators.

The probabilities associated with the numbers of winners per draw, and the possible lengths of droughts, are influenced by the numbers players tend to play. However, the sales data and number preferences needed to calculate these probabilities are rarely revealed by lottery operators. Here, we describe how attempts at estimating sales numbers are complicated by the non-uniformity in the numbers players play. We then use a simple but little-known method, based on the within-draw pattern of the numbers of prize-winners, to identify which numbers people tend to select or avoid. The numbers that players select are remarkably consistent across countries, and their non-uniformity within countries will continue to produce longer jackpot droughts than would be expected if the selections were entirely random or computer-generated.

The 2021 drought (62 consecutive rollovers of the jackpot prize) in the Irish Lottery saw some commentators blame the new structure, where players now have to match the six balls drawn out of 47, rather than out of 45 in the older version. One politician told the administrators that "the prospect of winning is [now] so remote that punters must be thinking [the racehorse] Shergar would have a better chance at winning Squid Game". Under the new structure, in the 601 draws up until the last pre-drought win in early June 2021, the jackpot had been won in 63 (10.5%) of them: the median drought was 8 draws, and the longest was 21. As the drought lengthened beyond 40, some politicians asked for an audit to reassure the public about the integrity of the lottery, and journalists began to ask statisticians what were the chances of this long a drought, and of it ending at the next draw.

To provide a first approximation to such probabilities, statisticians need to know how many tickets were bought for each previous draw, or how many are likely to be bought for the upcoming one. (Although players may make multiple selections or plays which appear as separate lines on the same ticket, we refer to these lines as tickets.) As with most lotteries, neither the Irish nor the UK lottery operator provides per-draw sales figures. Without such data, one can *estimate* the numbers of tickets sold. In some lotteries, the percentages of prize money allotted to various "tiers" are fixed. (Lottery administrators use the term "tiers" for the various prize categories, beginning with the jackpot prize for matching all numbers, the large prize for matching all but one number and a series of smaller prizes for matching fewer numbers.) When the percentages are fixed, it is possible to back-calculate the volume of sales, and thus the number of tickets, for each previous draw. Otherwise,

James Hanley is in the Department of Epidemiology, Biostatistics, and Occupational Health in the Faculty of Medicine at McGill University.

Michael Cronin is head of statistics in the School of Mathematical Sciences at University College Cork.

The belief that 7 is a lucky number is manifested in each of the countries we considered and the "unluckiness" of 13 can be seen in the UK data

one can estimate them from the number of prizes. For example, in the current Canada 6/49 lottery (i.e., six balls drawn out of 49), every ticket has a one in 6.62 chance of winning some kind of prize. Based on this, the estimated number of tickets sold for a past draw is 6.62 times the number of prizes won. However, because of human preferences in choosing numbers, and the resulting non-uniformity, the number of winners in draws with similar numbers of tickets sold varies widely. Recent data we have assembled from South Africa, Canada and the UK, where the numbers of selections are known, confirm this (bit.ly/3k5e2TB).

When there have been several rollovers, the expected returns from playing unpopular numbers are more attractive, so there is considerable interest in having good estimates of the popularity of the individual numbers. In a few isolated instances operators have provided authors with data on individual selections: 69 million tickets (9 million chosen by Lucky Dip) in one non-rollover week in 1996 in the UK lottery,[1] 2.6 million online transactions for the Dutch Lotto in 2010,[2] and 115 million manually chosen number sequences in all 118 draws of the Israeli national lottery in 2018.[3] Obviously, such data provide considerable detail, not only on how often different individual numbers are chosen by players, but also on combinations or sequences of numbers, such as 1, 2, 3, 4, 5, 6.

Without access to these privileged individual-selection data, statisticians have instead used publicly available lottery results and statistical models to learn the relative frequencies with which individual numbers are selected. Early on, the methods relied on the winning numbers drawn at each draw, along with the numbers of winners in some or all of the prize categories. Most methods also required either the actual or the estimated number of selections

Box 1: An example of the bonus-number estimator

The Irish Lotto has eight prize tiers: in four of them, a winner must match from 2 to 5 of the first six numbers drawn, as well as a final bonus number. In the other four tiers, the player must match from 3 to 6 of the first six numbers, but not the bonus number. With its 6/47 structure, out of the 10,737,573 possibilities, there are 6 + 600 + 15,600 + 148,200 = 164,406 combinations in tiers involving the bonus number and 1 + 240 + 11,700 + 197,600 = 209,541 in tiers that do not, a theoretical ratio of 0.785 : 1.

Over 565 draws, the overall ratio of the numbers of winners with the two types of matches was 8,612,208 : 11,008,172 or 0.782 : 1. In the subset of draws where the bonus number was 7, the ratio was 255,304 : 255,702, which is 28% higher than the overall ratio. If players were to select numbers randomly, the deviations from the overall ratio would vary randomly, with a small amplitude, around 0.
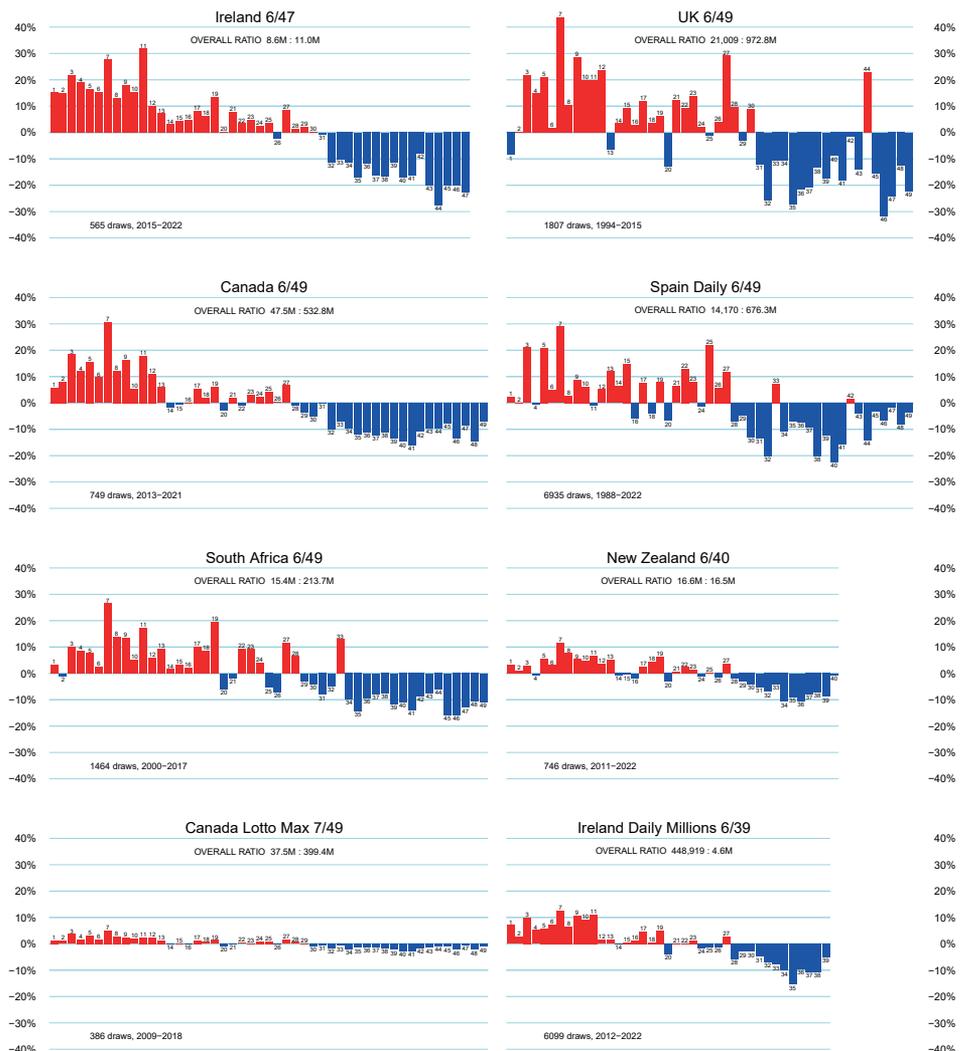

FIGURE 1: The extents to which the ratio of the numbers of winning selections in prize categories that do and do not involve a specific bonus number deviate from the overall ratio (see Box 1). If players were to select numbers blindly, the deviations from the overall ratio would vary randomly, with a very small amplitude, around 0. Instead, players who select their own numbers are more likely to select ones in the lower than the upper end of the range.

Box 2: Dealing with small counts

In lotteries where "match all but one plus the bonus" is the only tier that involves the bonus number, the small expected number of prize-winners in that tier meant that one particular draw could overly influence the estimated ratios. For example, in the UK lottery, where the mean number of winners in this tier was just 13.5 per draw, had we retained all 2,065 draws, the 50 draws in which the bonus number was 4 would have produced a ratio of 1.52. This extreme ratio was largely influenced by one draw, involving the winning numbers 12, 19, 27, 39, 42, and 47, where there were 243 winners in this tier. After we trimmed 10% of the draws, the ratio for the number 4 was 1.15.

In the Spanish lottery, in almost 8,000 draws, the mean number of winners in this tier was 15.9 per draw. The data from all 162 draws in which the bonus number was 23 would have produced a ratio of 2.17. This very extreme value was largely influenced by one draw, involving the winning numbers 22, 24, 25, 26, 27, and 38, where, thanks to the special position of the bonus number, there were 866 winners in this tier. After trimming, the ratio for the number 23 was 1.08.

played in each draw, and some required considerable computational time. Thus, for data from lotteries that have sizeable numbers of winners in tiers that involve an additional bonus number, the "bonus-number estimator"[4] (Box 1) was an important improvement. In addition to being easy to apply, this not very well-known method does not require knowledge of the numbers of selections played in each draw.

Lotteries in many countries now have bonus-number tiers with large numbers of winners, and their lottery results are publicly available (for example, via www.lotto.net and www.lottonumbers.com). We applied a modification of the bonus-number estimator to data from several of these. (Box 2 explains how we dealt with lotteries that did not have sizeable numbers of winners in tiers that involve an additional bonus number.)

The resulting estimated player preferences are shown in Figure 1. The preference for numbers in the lower end of the range confirms the patterns first seen in 1985 data from Canada (bit.ly/3vPvKx3), 1991–1993 data from California,[5] and 1994–1996 data from the UK.[6] Although we could not find any recent surveys on the reasons for this, the most commonly cited one has been the selection of numbers that match the birth months (1–12), and the birth days (1–31) of players and their families. The belief that 7 is a lucky number is manifested in each of the countries we considered, and the "unluckiness" of 13 can be seen in the UK data. Local peculiarities also appear; for example the relative popularity of 11, 19 and

27 in Ireland might be explained by the fact that these numbers appear in the rightmost column in the paper slip (bit.ly/3IzW6e1) a player fills in to play the lottery.

The near-uniformity in the Canadian (7/47) LottoMax is also instructive. It can be explained by the fact that the minimum purchase is $5 and comprises three sets of seven numbers; the player can either pick one set manually and get two additional computer-generated ("Quick Pick") sets, or can ask for three computer-generated sets. Thus, at least two-thirds of the lines played are selected randomly. Our own observations suggest that individuals and groups of players who play lotteries regularly tend to select their own numbers, while those who play casually, or only when the jackpot is large, are more willing to let the lottery computer select their numbers.

Incidentally, LottoMax is one of the few lotteries that discloses the per-draw numbers of tickets purchased for each draw. From these ticket-sales data (bit.ly/3k5e2TB), one can see the near-uniformity reflected in the smaller variations in the number of prizes in draws with similar-sized sales. In all but three of the LottoMax draws where we have ticket-sales data, the number of winners was within a few per cent of the expected number (1/19.8 times the number of lines played). But in three draws, the number of winners was *less than half* of what was expected. In one of these, the winning sequence was 2, 5, 6, 43, 46, 47, 48 (bonus 39); this unusual sequence (with three

numbers below 10 and the remainder all over 40) could only be identified as an outlier thanks to the availability of sales data. The winning sequences in the other two were 11, 18, 23, 24, 27, 28, 39 (bonus 1) and 1, 10, 12, 20, 29, 35, 47 (bonus 4).

Space does not permit us to address in any detail lotteries where *players themselves* select an additional (separate) supplementary number, to be matched with a supplementary number drawn from a second urn. Interested readers can consult separately the estimated preferences in the US Powerball and Mega Ball lotteries (bit.ly/3WOZ6aR) and the corresponding German, Spanish and French lotteries (bit.ly/3ilcXH5). The latter link also shows the estimated patterns in the EuroMillions lottery, where players select two extra numbers. Suffice it to say that, just as we saw above, there is "consistent non-uniformity".

Without full access to information,[4] it is not possible to calculate the probability that the jackpot will be won. However, the non-uniformity in the frequencies of the numbers selected by players already makes it more difficult for there to be a jackpot winner than if all the selections were random or computer-generated. We expect there will continue to be longer than expected jackpot droughts. Like climate change, portions of these long droughts are the result of human behaviours, rather than the physical behaviour of the lottery mechanism. ∎

References

1. Simon, J (1999) An analysis of the distribution of combinations chosen by UK national lottery players. *Journal of Risk and Uncertainty*, **17**(3), 243–276.
2. Wang, T. V., Potter van Loon, R. J. D., van den Assem, M. J. and van Dolder, D. (2016) Number preferences in lotteries. *Judgment and Decision Making*, **11**(3), 243–259.
3. Polin, B. A., Ben Isaac, E. and Aharon, I. (2021) Patterns in manually selected numbers in the Israeli lottery. *Judgment and Decision Making*, **16**(4), 1039–1059.
4. Roger, P. and Broihanne, M. (2007) Efficiency of betting markets and rationality of players: Evidence from the French 6/49 Lotto. *Journal of Applied Statistics*, **34**(6), 645–662.
5. Finkelstein, M. (1995) Estimating the frequency distribution of the numbers bet on the California lottery. *Applied Mathematics and Computation*, **69**, 195–207.
6. Farrell, L., Hartley, R., Lanot, G., Ian Walker, I. (2000) The demand for Lotto: The role of conscious selection. *Journal of Business & Economic Statistics*, 18(2), 228–241.