nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-020-01045-w

Check for updates

The association between gambling and financial, social and health outcomes in big financial data

Naomi Muggleton[1,2,3] ✉, Paula Parpart[2,3,4], Philip Newall[5,6], David Leake[3], John Gathergood[7] and Neil Stewart[2]

Gambling is an ordinary pastime for some people, but is associated with addiction and harmful outcomes for others. Evidence of these harms is limited to small-sample, cross-sectional self-reports, such as prevalence surveys. We examine the association between gambling as a proportion of monthly income and 31 financial, social and health outcomes using anonymous data provided by a UK retail bank, aggregated for up to 6.5 million individuals over up to 7 years. Gambling is associated with higher financial distress and lower financial inclusion and planning, and with negative lifestyle, health, well-being and leisure outcomes. Gambling is associated with higher rates of future unemployment and physical disability and, at the highest levels, with substantially increased mortality. Gambling is persistent over time, growing over the sample period, and has higher negative associations among the heaviest gamblers. Our findings inform the debate over the relationship between gambling and life experiences across the population.

Gambling has existed for millennia in a variety of forms[1]. New gambling markets continue to emerge in many countries, including the United States, where a recent Supreme Court ruling deemed sports betting to be legal in every state[2]. In 2019, in the United Kingdom (the context for this study) over 24 million individuals collectively lost over £14.5 billion to bookmakers, casinos, lotteries and other gambling platforms[3]. The UK public's gambling losses have steadily increased over recent years, as mobile and online technologies make gambling more available than ever before[4]. Advertising has also increased the visibility of gambling since 2007[5] with, for example, one in six adverts shown during the broadcaster ITV's programming for the 2018 FIFA World Cup promoting gambling[6]—an event that led to calls from some community and policy leaders for greater regulation[7]. This is an example of what some public health researchers have called the 'gamblification of sport'[8]. Yet the scientific and policy communities have highlighted the lack of reliable data available and the need for studies that examine the association between gambling and personal outcomes, including lifestyle and well-being, using objective data[9–12].

We analyse gambling behaviour via detailed, anonymous, individual-level financial transaction data from millions of customers of the United Kingdom's largest retail bank, Lloyds Banking Group (LBG). Our largest dataset tracks ~6.5 million people or around 10.6% of the population of the United Kingdom, over 7 years. Big financial transaction data provide a unique view of individual-level gambling behaviour, consisting of the full spread of electronic payments to gambling platforms, which allows us to identify the distribution (who, when and for how long) of gambling and its associated outcomes across a national population. The relationship of gambling with financial outcomes (for example, savings and debt) and non-financial outcomes (for example, spending on hobbies, social activities and night-time online spending), can all be inferred objectively and analysed alongside information on gambling behaviour. We also measure longer-term outcomes, including transitions into unemployment, disability and mortality. This view of individual outcomes is rivalled only by what a state monopolist could see—it cannot be seen in the data of gambling firms, self-reported survey data or the aggregated data reported by firms, industry groups and regulators.

This observational study documents gambling in the United Kingdom with large-scale objective data. Previous approaches had to rely primarily on self-report surveys and smaller sample sizes[13]. For example, the United Kingdom ran three waves of the British Gambling Prevalence Survey in 1999, 2007 and 2010, considered by expert witnesses in a recent government select committee as the best national data on gambling in the United Kingdom[14]. The 2010 survey used a sample of 7,756 respondents or ~0.01% of the then population of the United Kingdom[15]. This survey estimated that 0.7–0.9% of the then population of the United Kingdom met diagnostic criteria for disordered gambling, although this estimate is based on less than 100 cases, as is typical in prevalence surveys given population base rates[16]. It has been argued that these base rates may be understated if gamblers hide or cover-up their gambling when filling out these surveys[17]. Prevalence surveys also ask respondents to self-report their gambling involvement and expenditure. However, it has been demonstrated that disordered gamblers cannot self-report their gambling expenditure reliably[18], that memory biases are an established feature of disordered gambling[19] and that prevalence surveys may struggle to recruit sufficient disordered gamblers given population base rates[16]. Similar[20], or smaller sample sizes[21,22], have so far been used to examine the relationship between gambling and mortality. A further advantage is that transaction data take the form of individual-level panels which follow the same individual over time. To date, most gambling research is cross-sectional in nature, with a comparative lack of longitudinal studies[23]—which exhibit increased levels of attrition amongst disordered gamblers[24]. By comparison, our big financial transaction data approach

[1]Department of Social Policy and Intervention, University of Oxford, Oxford, UK. [2]Warwick Business School, University of Warwick, Coventry, UK. [3]Applied Science, Lloyds Banking Group, London, UK. [4]Department of Experimental Psychology, University of Oxford, Oxford, UK. [5]Warwick Manufacturing Group, University of Warwick, Coventry, UK. [6]School of Health, Medical and Applied Sciences, CQ University, Melbourne, Australia. [7]School of Economics, University of Nottingham, Nottingham, UK. ✉e-mail: naomi.muggleton@spi.ox.ac.uk

unobtrusively follows a random sample drawn from a substantial fraction of the banked population of the United Kingdom.

The empirical gambling-related harm literature has added a focus on the negative consequences associated with gambling but is also limited by a focus on cross-sectional self-report surveys[25–28]. Thus far, there have been two main attempts to create conceptual frameworks to better understand the multidimensional nature of the relationship between gambling and individual outcomes[29,30]. Langham et al.[29] derived a list of 72 distinct 'harms', covering financial, relationship, psychological, health, work, study and social deviance harms. Later research has shown that these harms differ markedly with respect to prevalence, with financial harms being the most prevalent and social deviance harms the least prevalent[31]. Wardle et al.[30] conceptualized gambling-related harm as affecting economic resources, relationships and health, with harms potentially having persistent effects through time, and being felt beyond individuals and across wider communities. Moreover, there is a current debate about the extent to which gambling harms are concentrated amongst disordered gamblers[32,33] versus the overall impact of harm felt amongst the larger group of lower-risk gamblers[31].

We contribute to this literature with a data-driven approach. Our analysis focuses on quantifying the association between gambling and personal outcomes. The evidence we present raises questions of causation and the mechanisms by which associations arise, which are topics for future work.

Results

Levels of gambling. We used a random sample (sample 1) of 102,195 customers active in each month of 2018. The unit of analysis in this panel data sample is an account calendar month. To identify gambling transactions, we relied on the pre-existing gambling category in the bank's typography of transactions, which includes various forms of gambling such as offline and online bookmakers, casinos, lotteries and other providers. Cash gambling and gambling at other types of retailers (for example, a lottery ticket at the supermarket) are not captured and thus we are conservative in estimating total gambling.

Summary data in Table 1 reveal that 43% of individuals in the sample made at least one electronic gambling transaction in 2018. Among those who made at least one electronic transaction, the median number of transactions was 12 (mean = 56), with a median year spend of £125 (mean = £1,345), which is approximately a median of 0.5% of monthly spending (mean = 4%). The gap between the mean and median values highlights the highly skewed nature of gambling behaviour (Supplementary Table 1). We define spend as the sum of all gambling transactions that were processed via a debit card or credit card. The distribution of spending has a long right-tail, with the top 10% of gamblers spending >£1,800 on gambling in the calendar year, close to 8% of their total spending.

Gambling and financial stress. Here, we describe how gambling is associated with financial distress, financial inclusion and financial planning in a random sample of active customers (sample 1) (top rows of Fig. 1). The unit of analysis in this sample is a calendar month. The measures of financial distress are: using an unplanned overdraft, missing a credit card payment, taking a payday loan, missing a loan repayment and missing a mortgage repayment. Financial inclusion measures are: having a credit card, loan or mortgage, credit card use and making a payment to a debt recovery agency. Financial planning measures are: holding insurance, paying down a mortgage, saving money, saving money in a tax-preferred savings account (known as an individual savings account (ISA) in the United Kingdom) and paying into a self-invested pension. A detailed description of all the outcome variables is contained in Supplementary Table 2, with summary statistics reported in Supplementary Table 1. The set of outcome variables shown includes all outcomes that were analysed.

In all of the binned scatterplots related to financial outcomes in Fig. 1 (rows 1–3) the unit of analysis is one account calendar month. For each account month, we calculated the percentage of the individual's total spend in that month devoted to gambling. Total spend was calculated by summing all outflows of cash across a given month, and included credit card, debit card, direct debit and ATM transactions but not internal movements of money (for example, movement from a personal current account to savings account). The *x* axis shows the percentile rank of this variable. Each panel contains 101 dots. The dot at 0% on the *x* axis include account months in which the individual had zero gambling (not all individuals who gamble do so in each month of the sample period). That is, if a gambler had an account month where they did not gamble, he or she would be captured in the dot at 0%. Each of the remaining 100 dots represent one percentile of account months (typically 150–3,000 account months, depending on the sample size; Supplementary Tables 4–6). Thus, the dot at 1% represents the 1% of observations where gambling was lowest (but not zero) and the dot at 100% represents the 1% of observations where gambling was highest. (The discontinuity between 0% and 1% results for technical reasons: selecting accounts with zero gambling selects accounts that were less likely to be active for other transactions.) The *y* axis shows the mean value of the dependent variable at each percentile. For this analysis, the dependent variable is measured 1 month forward, to avoid a mechanical relationship whereby higher gambling mechanistically reduces the value of outcome variables related to spending due to individuals having less net income to spend on other items in months when more is spent on gambling. The lines are penalized cubic regression splines estimated directly from the underlying data with 95% confidence intervals (CI).

Higher gambling is associated with a higher rate of using an unplanned bank overdraft, missing a credit card, loan or mortgage payment, and taking a payday loan. A 10% point increase in absolute gambling spend is associated with an increase in payday loan uptake by 51.5% (so, for example, 0.97% of those with 0% of spending on gambling have a payday loan but 1.47% of those with 10% of spending on gambling have a payday loan, an increase of 51.5%) and the likelihood of missing a mortgage payment by 97.5% (Supplementary Table 3). In all reported cases, the effect of a 10% point increase in absolute gambling spend are reported after controlling for age, gender and annual income.

Gambling is associated with lower rates of holding a credit card, loan or mortgage, higher use of credit card balances and a higher likelihood of the individual being subject to debt collection by bailiffs. A 10% point increase in absolute gambling is associated with an increase in credit card use by 11.2% and bailiff interaction by 8% (Supplementary Table 3). Conversely, higher gambling is associated with smaller spends on insurance and mortgage repayments, smaller total savings and smaller pension contributions. For many of the outcome variables, the association with gambling is notably stronger at high percentile ranks approximately above the 75th percentile (which equates to ~3.6% of total monthly expenditure). This suggests that the relationship between gambling and many of the harmful outcomes is stronger when the individual is devoting a relatively large share of total monthly spending to gambling.

We conducted regression analyses, using an ordinary least squares regression estimator in a specification that controlled for age, gender and income in addition to gambling as a percentage of monthly spend (all variables entering linearly, together with a constant term). All statistical tests were two-sided. The coefficients on the gambling covariates, together with 95% CI and marginal R^2, are reported in Supplementary Table 3 (with the full regression estimates reported in Supplementary Tables 4–6).

Table 1 | Summary statistics for sample 1

	Mean	s.d.	Percentiles				
			25th	50th	75th	90th	99th
Panel A: individual annual totals							
Gambling transaction in 2018 (1/0)	0.43						
Number of transactions	24.31	118.37	0	0	10	35	515
Number of transactions (>0)	56.05	174.74	3	12	30	112	843
Transactions (£)	583.30	8,907.18	0	0	110	498	11,200
Transactions (£, >0)	1,345.17	13,488.58	40	125	438	1,831	22,060
Transactions as percentage of spending	1.59	7.02	0	0	0.39	1.92	40.12
Transactions as percentage of spending (>0)	3.67	10.30	0.17	0.53	1.70	7.91	58.18
n	102,195						
Panel B: individual × months							
Gambling transaction in month (1/0)	0.26						
Number of transactions	2.05	11.16	0	0	1	3	45
Number of transactions (>0)	7.79	20.71	1	2	5	17	100
Transactions (£)	49.17	911.22	0	0	8	40	908
Transactions (£, >0)	186.83	1,769.01	10	22.50	70	275	2,723.70
Transactions as percentage of spending	1.53	7.28	0	0	0.22	1.85	40.87
Transactions as percentage of spending (>0)	5.83	13.27	0.54	1.16	3.64	15.11	71.94
n	1,210,632						

Gambling, lifestyle and well-being. Outcomes associated with gambling extend beyond the purely financial (bottom rows of Fig. 1). The wider themes are lifestyle (spend on fast food, gaming, bars, tobacco and off licences), health and well-being (spend on prescriptions, self-care, fitness and night-time spending between 1:00 and 5:00) and leisure and interests (spend on hobbies, social activities, education and travel), which are analysed in a random sample of active customers (sample 1), where the unit of analysis is a calendar month. Results show a negative association between gambling and self-care, fitness activities (for example, gym membership), social activities, and spending on education and hobbies. There is also an association between gambling, social isolation and night-time wakefulness—individuals spending more on gambling travel less and are more likely to spend at night. A 10% point increase in absolute gambling equates to an 11.5% increase in nights awake and 9% reduction in social activities (Supplementary Table 3). The relationship between gambling on reduced socialization is also seen in lower spend at bars and pubs. But higher levels of gambling are associated with lower off-licence spending. The relation with fast-food spend is more complex (see Supplementary Table 3 for regression coefficients, with the full regression estimates reported in Supplementary Tables 7–9). The coefficient estimates are precisely estimated and confirm the directional relations illustrated in Fig. 1, with the exception being tobacco spend, for which the coefficient is not precisely estimated.

Gambling, unemployment, disability and mortality. Here, we describe medium-term associations with unemployment, disability and mortality using data from all 6.5 million active customers in each month in 2013 (sample 2). We tracked these individuals across the subsequent 5 years, 2014–2019. We find that higher gambling is associated with a higher risk of future unemployment and future physical disability. The panel 'Disability payments' in Fig. 1 restricts sample 2 to individuals who were not receiving disability payments in 2013 and plots the relationship between the percentile rank of gambling spend as a percentage of monthly income and the likelihood of subsequently claiming disability payments over the period January 2014 to July 2019. The plot reveals a positive association (Supplementary Table 10).

The panel 'Unemployment' in Fig. 1 restricts sample 2 to individuals who were employed in 2013 and plots the relationship between the percentile rank of gambling spend as a percentage of monthly income and the likelihood of subsequently experiencing at least one spell of unemployment over the period January 2014 to July 2019. The positive relationship is notably stronger at high levels of gambling, with employed individuals in the highest percentiles of gambling having a 6% likelihood of experiencing future unemployment (Supplementary Table 10).

We examined the relationship between gambling spend and mortality. We model mortality using survival analysis in adult males and females drawn from sample 2. We fitted Cox proportional hazard models to the data, controlling for amount gambled, individual gender and individual age. The model censors individuals who left the sample for reasons other than mortality. Figure 2 plots the Cox model fits, showing the relationship between levels of gambling, where gambling is expressed as a proportion of monthly income of 0%, 10%, 20% or 30%. (Table 1 shows that the top 1% of gamblers gambled over 58% of their income in 2018.) The *x* axis plots time in years (from January 2014) and the *y* axis plots the survival probability. Plots are shown for men and women at three age points. For all groups, the survival probability is lower at higher levels of gambling. Information is not available on cause of mortality. The heaviest gamblers exhibit higher 5-year mortality rates. For example, among 44-year-old women, gambling 30% of annual expenditure (relative to 0%) is associated with an increased chance of death from 50 in 10,000 (95% CIs [50, 51]) to 69 in 10,000 (95% CIs [66, 72]) or by a factor of 1.37 (Supplementary Table 11). High levels of gambling are associated with a likelihood of mortality that is about one-third higher, for both men and women, younger and older.

The time course of gambling. Gambling is also persistent over time, although individuals can transition into (and out of) high levels of gambling within a few months. We used a random sample of 101,151 customers active over all months from 2012 to 2018

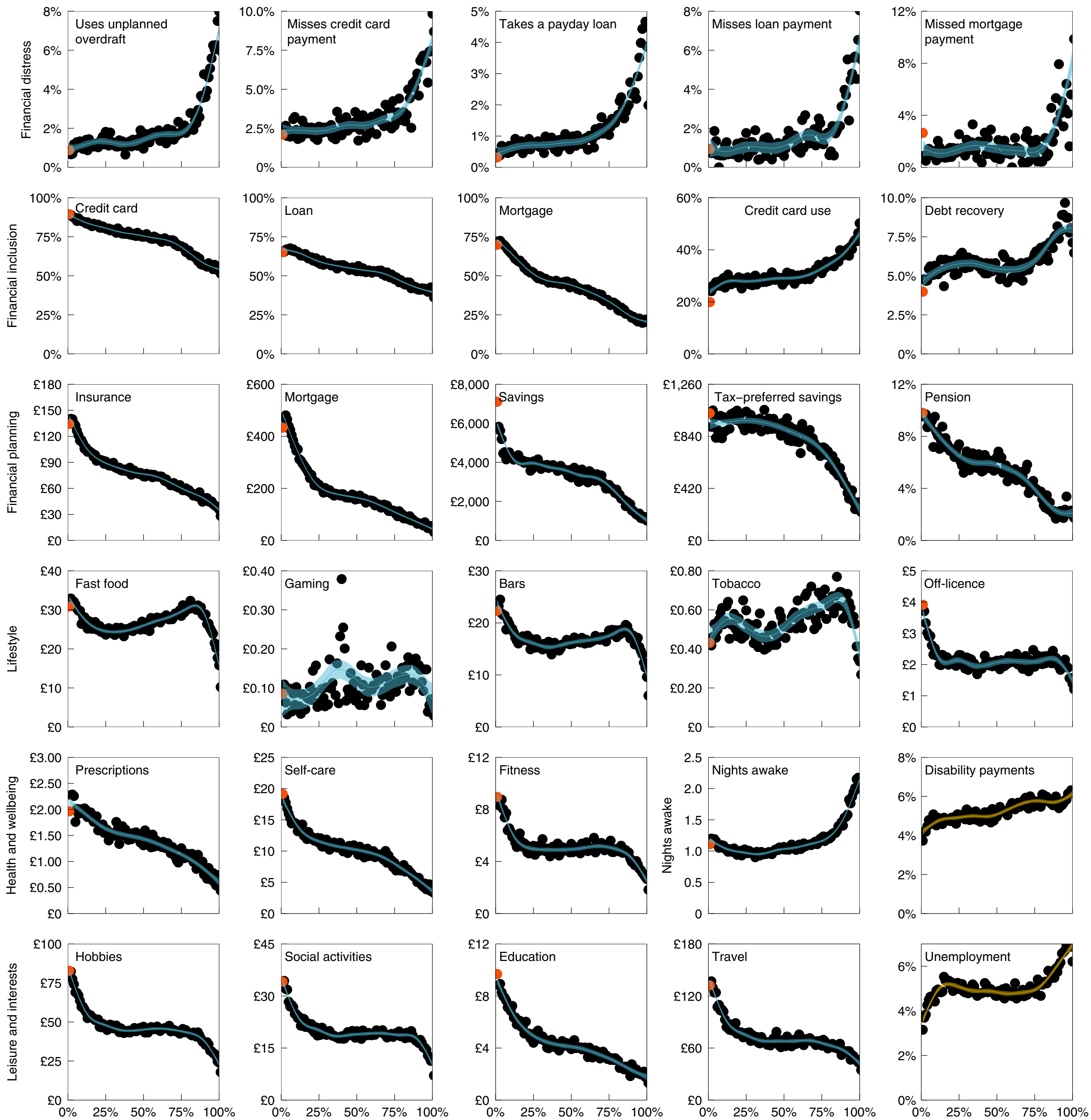

Fig. 1 | Gambling and financial, lifestyle and well-being outcomes. Binned scatterplots with account months binned by their gambling percentile rank at $t=0$ on the x axis. The sample is restricted to individuals who gambled at least once in 2018, so is not generalizable to those who did not gamble during the observation period. Individuals who did not gamble in $t=0$ but did gamble at some point in 2018 are captured at 0% (red dot). Account months with gambling are binned into 1% bins on the basis of the percentage of the total spend gambled in a month. Means of the dependent variable at $t=1$ for each bin are plotted on the y axis. The trend line shows smoothing with cubic regression splines on the underlying raw data. Shading denotes 95% CI. Financial distress measures are: probability of entering an unplanned overdraft; missing a debt repayment for credit cards, loans or mortgages; and taking a payday loan. Financial inclusion measure are: having a credit card, having a loan, having a mortgage, credit card use, and making a debt recovery payment. Financial planning measures are: holding insurance, paying down mortgage, saving money, tax-preferred account saving and self-invested pension saving. Lifestyle, health and well-being, and leisure and interests outcomes are measured in £; with the exceptions of disability payment receipt and unemployment, which are measured as a percentage of the sample. All blue plots are based on estimates for sample 1 ($n=102{,}195$) and orange plots are based on estimates for sample 2 ($n=6{,}515{,}557$).

(sample 3). The top panel of Fig. 3 illustrates the movement over time of individuals between levels of gambling. The analysis is centred on the year 2015, showing the level of gambling that leads to and leads from 2015. Gambling is persistent but some small fractions of individuals move from no gambling in 2012 to the highest levels in 2015 and some small fractions gambling at the highest level

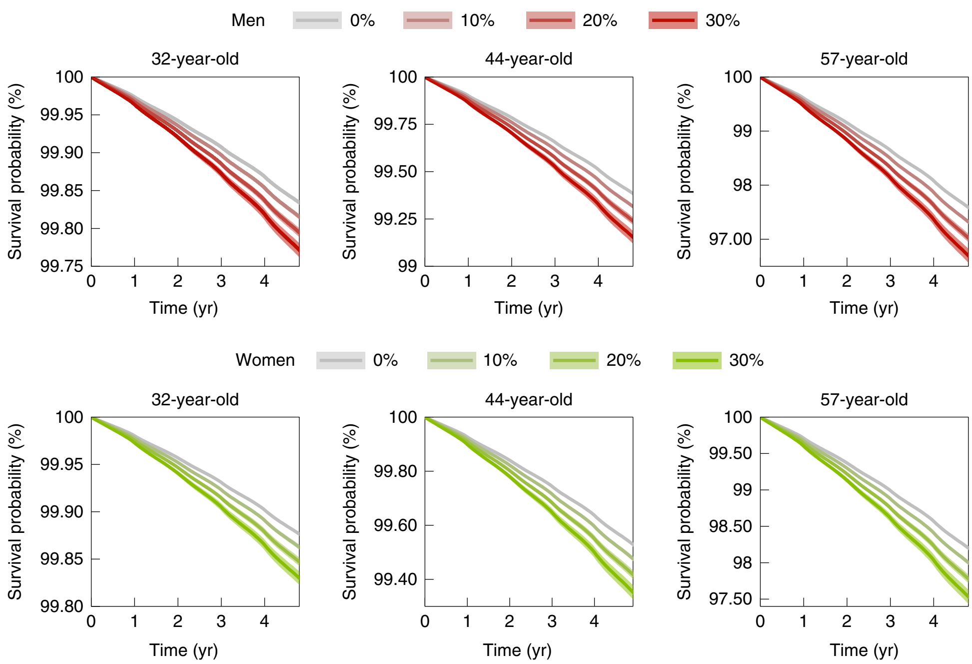

Fig. 2 | Gambling and mortality. Cox regression survival analysis of mortality rate (time in years) beginning in 2014 ($n = 6,515,557$). Survival is modelled as the percentage of total spend gambled in 2013 (colour intensity) while controlling for gender and age in 2013 (colour and panel). The model censors individuals who left the sample for reasons other than mortality (for example, switched bank). Shading denotes 95% CI. Note that the *y* scale varies between panels.

in 2015 have stopped in 2018. The bottom panel zooms in on the highest-spending gamblers to see whether they have always gambled heavily in the past. The sample comprises a subset from sample 3 whose gambling was >10% of their total spending in Quarter 2 of 2015 (2,168 individuals). We find that, for example, 3 years earlier around half of the highest-spending gamblers were already gambling heavily, while only 6 months before, over 6.9% of these heavy gamblers were not gambling at all, highlighting the fast acceleration with which some individuals can transition into heavy gambling. In contrast, 6 months later 4.6% of heavy gamblers were not gambling at all. This asymmetry shows that gambling expenditure represents sticky behaviour.

Discussion

This paper demonstrates that financial transaction data can produce a view of gambling-related outcomes that is objective, longitudinal and mass-scale. By comparison, prevalence surveys, which have dominated the view that academics and policy-makers have of gambling for the last 30 years, are self-report, cross-sectional and largely small sample in nature[13]. We described the association between gambling and 31 outcome variables from the financial and wider social and health domains. Given that our data do not cover cash gambling transactions, or electronic transactions using third-party payment processors or another person's account details, the estimated effects of gambling expenditure on gambling-related harm are probably conservative. Our evidence complements existing approaches, which draw upon self-report surveys, case studies or inferences from industry or aggregate-level statistics[13,21,22,25–28,34–36] by relying on large-scale objective data. As such, the reported findings have implications for the future study of gambling epidemiology and public health.

This study contains some limitations that could be addressed with future research. First, and similarly to gambling prevalence surveys, we do not establish causality, which means that findings demonstrate associations that may reflect causality or comorbidity—both of which are of concern. Causality would indicate that higher levels of gambling increase one's risk of negative outcomes like financial distress, social exclusion, disability and unemployment. Comorbidity, however, would indicate that individuals who are susceptible to these negative outcomes due to other factors are more likely to be drawn to gambling. In reality, the observed effects could result from a blend of causality and comorbidity, both of which have significant implications for policy-makers and public health experts. Further work is needed to measure the extent to which gambling-related harm is driven by causal mechanisms and/or whether gambling firms increasingly target the most vulnerable members of society through advertising and the selection of store locations. Second, our methodology does not rule out the possibility of reverse causation, such that an increase in harm precedes an increase in gambling. To partially overcome this, we use measures of gambling at t_0 to predict outcomes at t_1 to exclude scenarios where, say, missing a credit card payment leads to an individual gambling as a means to pay off debt. Yet, as we have shown, gambling is highly persistent across time. As such, it is possible that gambling may co-occur, or be preceded by, negative life events. Third, we are unable to extend our analysis beyond a 6-year window of transactional data. It is possible that the breadth of harms associated with gambling, such as mortality, disability or unemployment, might look different when analysed across a longer period of time. Fourth, the breadth of our analyses means that we cannot control for all social, economic and political events that occurred in the 2012–2018 window of our study. Finally, our analyses were conducted among a sample of banked UK residents. Further work is needed to test the generalizability of our findings among other populations.

Nonetheless, a longitudinal financial transaction approach informs the current gambling policy debate. Some argue that associations between gambling and negative outcomes exist primarily among a small group of disordered gamblers, who should be the focus of mitigating gambling-related harm[32,33,37]. In support of this view, we find a number of negative outcomes such as nights

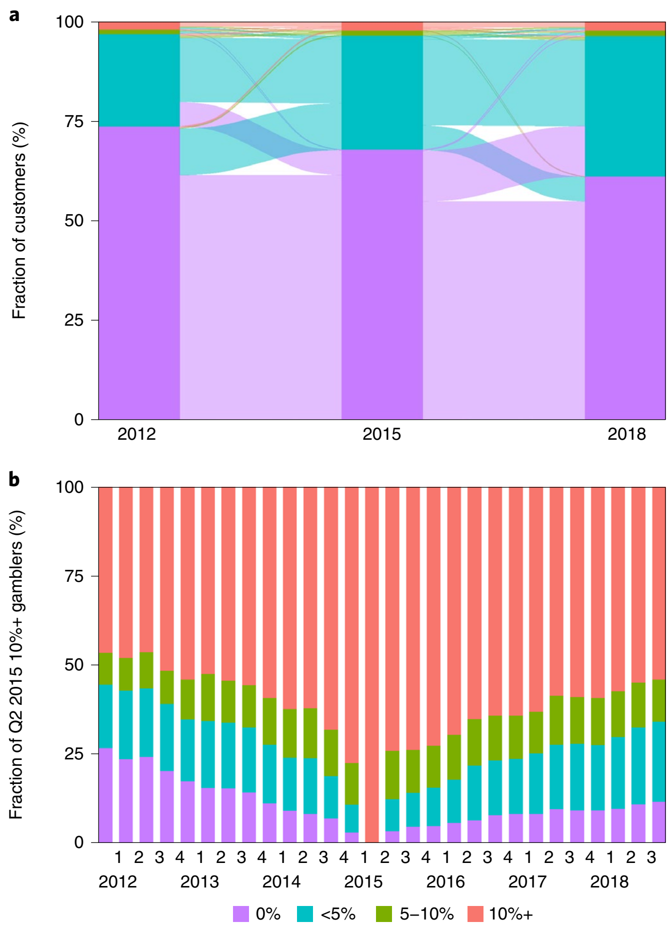

Fig. 3 | Persistence of gambling. The movement of individuals between levels of the percentage of total spend gambled ($n = 101{,}151$). **a**, Flows of individuals into and out of levels of gambling in 2015. **b**, Levels of gambling over quarters in a subset of individuals gambling >10% of their total spend in Quarter 2 of 2015 ($n = 2{,}168$).

awake, unemployment and mortality, that increase markedly for the highest-spending gamblers. By comparison, others argue that the share of the population experiencing significant gambling-associated harms is broader than this small group of disordered gamblers[26,31,34] and that policy should be similarly broad-based. In support of that view, we find that more gambling is associated with more negative outcomes even at lower levels of gambling, and that individuals can rapidly transition between different levels of gambling. Overall, our findings suggest that policy-makers may want to do more to efficiently detect and protect the highest-spending gamblers, while also attempting to control population levels of risk[38].

Methods

Ethical approval. The Privacy Risk and Impact Assessment Committee at LBG granted ethical approval for this study on aggregated, anonymous data as part of a strategy to help vulnerable customers. Upon opening an account, LBG customers consented for their data to be used for research: https://www.lloydsbank.com/help-guidance/customer-support/privacy-explained/data-privacy-notice.html. The Humanities and Social Sciences Research Ethics Committee at the University of Warwick waived the requirement for an additional ethics review, as in cases where appropriate ethical review has already taken place at another collaborating institution, Warwick ethical review is not required so as to avoid unnecessary duplication.

Sample selection. Our sample contains a large subsection of the banking population of the United Kingdom. Of the 52.4 million adults in the United Kingdom, 1.5 million (2.9%) were unbanked[39]. Of the total adult population, our in-scope sample was ~10.6% of the adult UK population. Owing to the diversity of outcome variables and time frames analysed above, we required three distinct samples. We used LBG's definition of an active customer as an individual whose account(s) process at least 12 transactions per month. This definition was constructed independently of the authors and before the analysis commencing. It avoids including cases where individuals hold dormant bank accounts. The inclusion criteria also ensured that all individuals were aged 18 years or older (the legal age for gambling in the United Kingdom) during the observation time frame.

Multiple individuals can be assigned to the same bank account. But, for each account, we identify the primary account holder. We then source the transactions associated with that individual's debit card(s) and/or credit card(s). This means that, for a couple who share a bank account, only the primary account holder would be eligible for inclusion in our sample selection and only transactions enacted on his or her debit/credit card(s) would be tagged to the associated account. As such, whereas all transactions for a joint account appear together on a statement, within the data we can assign transactions to the individual who initiated it because transactions are marked with a card identifier (unique to the individual) as well as an account identifier.

Sample 1 consists of a random sample of all individuals who were active each month throughout 2018. In this sample, we required that individuals were aged ≥18 years at the beginning of 2018. Thus, sample 1 is a random sample of individuals who held an active current account for each month in the calendar year 2018. Of the 5,394,933 individuals who met this criteria, we randomly selected 1/53 of customers, giving us a sample of 102,195 individuals. The unit of analysis in this sample is an account calendar month. Gambling behaviour was measured 1 month back to avoid a mechanical relationship between higher gambling spend and lower spend on other items, for a fixed monthly budget.

Sample 2 consists of a larger sample of all individuals to be able to detect comparatively rarer events over a 6-year time frame. This period was a time of relative stability in the United Kingdom, with no periods of economic recession or public health concerns. As such, we are confident that our analyses of mortality, unemployment and disability are generalizable and not artefacts of the observation period. Our analysis could not be extended beyond this time frame, as some sensitive data are deleted by the bank beyond this window. To ensure that we were not capturing dormant accounts, we required a sample of individuals who were active in 2013. But we did not select on account activity during our outcome window of 2014–2019, to allow for detection of our outcome variable (mortality) and control variable (individual leaving the bank). In this sample, we required that individuals were aged ≥18 years at the beginning of 2013. Sample 2 represents all individuals who held an active current account for each month in the calendar year 2013. The sample consisted of 6,515,557 individuals who were subsequently tracked between January 2014 and December 2018. The unit of analysis in this sample is an account calendar year. Gambling behaviour was measured and aggregated across 2013.

Sample 3 consists of a random sample of all individuals who were active each month throughout 2012–2018. In this sample, we required that all individuals were aged ≥18 years at the beginning of 2012. Hence, sample 3 is a random sample of individuals who held an active current account for each month from January 2012 to December 2018. Of the 5,281,778 customers who met this criteria, we randomly selected 1/52 of customers, giving us a sample of 101,151 individuals. The unit of analysis in this sample is account calendar month.

Measuring gambling behaviour. Gambling is measured by electronic transactions to gambling licensed firms identified by the bank in its typology of transactions. A transaction is defined as any spending behaviour that occurs using a debit card or credit card. This includes electronic transfers to gambling platforms, online gambling transactions, and chip and pin or contactless in-store transactions but neither cash transactions nor cheques. This was constructed independently of the authors and before the analysis commencing (the gambling category includes offline and online bookmakers, casinos, lotteries and other providers). This measure underestimates total gambling, as it does not include cash gambling and transactions where gambling might occur through a general retailer (such as lottery tickets purchased as part of a supermarket shop). It also omits gambling in cases where an intermediate transaction to a payments platform (for example, PayPal) is used to make a subsequent gambling transaction.

Variable construction. Our variables are a combination of account status flags within LBG (for example, credit card arrears), sums over pre-existing categorizations of merchant transaction strings constructed independently by LBG (for example, spending on fast food) or from transaction metadata (for example, night-time expenditure inferred from time stamps on manual transactions). Our definition of transaction is the same as that outlined in the previous section (Measuring gambling behaviour). A detailed description of all the outcome variables is contained in Supplementary Table 2, with summary statistics reported in Supplementary Table 1. Data distribution was approximately normally distributed but this was not formally tested.

The set of outcome variables shown includes all outcomes that were analysed. In addition to those shown, we attempted to build the following measures, which could not be constructed and were therefore not analysed:

1. Divorce: infeasible given the limited information on marital status that can be inferred from transaction records
2. Health spending: infeasible due to ambiguity over the purpose of specific health spends (for example, distinguishing preventative health care spending from treatment costs). Therefore, we created the more clearly interpretable 'self-care' measure (Fig. 1)
3. Number of public transport transactions: infeasible due to the ambiguity over interpretation of public transport spend. For example, whereas the frequency of public transport transactions may correspond to higher mobility in cities, it could also be a sign of poverty in rural areas
4. Hospital spend, number of national health service (NHS) visits, rent spend and estate agent spend: infeasible due to the limited number of transactions that could be classified as such from transaction strings

Robustness checks. *Replication with only gamblers.* The regression analyses conducted in Supplementary Tables 4–10 are carried out on all individuals and are not contingent upon whether they gambled during the observation period. But this raises important questions regarding the generalizability of our findings. As a robustness check we have replicated the analyses outlined in Supplementary Tables 4–10 in Supplementary Tables 12–18. Here, we show that, of our 30 outcome variables, 28 findings are replicated among the only-gamblers sample. The two exceptions are:

1. Gaming: inconclusive coefficient estimate in the full sample, $B=0.017$ 95% CIs [-0.059, 0.094], $P=0.658$; Supplementary Table 7) but negative coefficient estimate in the only-gamblers sample ($B=0.090$ 95% CIs [0.016, 0.013], $P=0.021$; Supplementary Table 15)
2. Tobacco: positive coefficient estimate in the full sample ($B=0.35$ 95% CIs [0.030, 0.68], $P=0.032$; Supplementary Table 7) but inconclusive coefficient estimate in the only-gamblers sample ($B=-0.35$ 95% CIs [-0.77, 0.056], $P=0.090$; Supplementary Table 15)

Replication controlling for seasonal effects. The unit of analysis in Supplementary Tables 12–18 is one calendar month. To control for the possibility of unaccounted-for associations between calendar months within individuals, we re-ran the analysis, adding clustered standard errors about the observation month (Supplementary Tables 19–24). Here, we show that, of our 28 monthly outcome variables, all 28 replicate the findings observed in the main analyses.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The data that support the findings of this study are available from LBG but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are available from the authors upon reasonable request and with permission of LBG.

Code availability
Data were extracted from LBG databases using Teradata SQL Assistant (v.15.10.1.9). Data analysis was conducted using R (v.3.4.4). The SQL code that supports the analysis is commercially sensitive and is therefore not publicly available. The code is available from the authors upon reasonable request and with permission of LBG. The R code that supports this analysis can be found at github.com/nmuggleton/gambling_related_harm. Commercially sensitive code has been redacted. This should not affect the interpretability of the code.

Received: 26 March 2020; Accepted: 23 December 2020;
Published online: 4 February 2021

References
1. Schwartz, D. G. *Roll the Bones: The History of Gambling* (Winchester Books, 2013).
2. Supreme Court of the United States. *Murphy, Governor of New Jersey et al. versus National Collegiate Athletic Association et al.* No. 138, 1461 (2018).
3. *Industry Statistics* (Gambling Commission, 2019); https://www.gamblingcommission.gov.uk/PDF/survey-data/Gambling-industry-statistics.pdf
4. Cassidy, R. *Vicious Games: Capitalism and Gambling* (Pluto Press, 2020).
5. Orford, J. *The Gambling Establishment: Challenging the Power of the Modern Gambling Industry and its Allies* (Routledge, 2019).
6. Duncan, P., Davies, R. & Sweney, M. Children 'bombarded' with betting adverts during World Cup. *The Guardian* https://www.theguardian.com/media/2018/jul/15/children-bombarded-with-betting-adverts-during-world-cup (15 July 2018).
7. Hymas, C. Church of England backs ban on gambling adverts during live sporting events. *The Telegraph* https://www.telegraph.co.uk/news/2018/10/08/church-england-backs-ban-gambling-adverts-live-sporting-events/ (10 August 2018).
8. McGee, D. On the normalisation of online sports gambling among young adult men in the UK: a public health perspective. *Public Health* **184**, 89–94 (2020).
9. Editorial. Science has a gambling problem. *Nature* **553**, 379 (2018).
10. *National Gambling Strategy* (Gambling Commission, 2019); https://www.gamblingcommission.gov.uk/news-action-and-statistics/News/gambling-commission-launches-new-national-strategy-to-reduce-gambling-harms
11. van Schalkwyk, M. C. I., Cassidy, R., McKee, M. & Petticrew, M. Gambling control: in support of a public health response to gambling. *The Lancet* **393**, 1680–1681 (2019).
12. Wardle, H., Reith, G., Langham, E. & Rogers, R. D. Gambling and public health: we need policy action to prevent harm. *BMJ* **365**, l1807 (2019).
13. Volberg, R. A. Fifteen years of problem gambling prevalence research: what do we know? Where do we go? *J. Gambl. Issues* https://doi.org/10.4309/jgi.2004.10.12 (2004).
14. Select Committee on the Social and Economic Impact of the Gambling Industry. *Corrected Oral Evidence: Social and Economic Impact of the Gambling Industry* (UK Parliament, 2019); https://committees.parliament.uk/oralevidence/15/html/
15. Wardle, H. et al. *British Gambling Prevalence Survey 2010* (National Centre for Social Research, 2011); https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/243515/9780108509636.pdf
16. Smith, G., Hodgins, D. & Williams, R. *Research and Measurement Issues in Gambling Studies* (Emerald, 2007).
17. Harrison, G. W., Lau, M. I. & Ross, D. The risk of gambling problems in the general population: a reconsideration. *J. Gambl. Stud.***36**, 1133–1159 (2020).
18. Wood, R. T. & Williams, R. J. 'How much money do you spend on gambling?' The comparative validity of question wordings used to assess gambling expenditure. *Int. J. Soc. Res. Methodol.* **10**, 63–77 (2007).
19. Toneatto, T., Blitz-Miller, T., Calderwood, K., Dragonetti, R. & Tsanos, A. Cognitive distortions in heavy gambling. *J. Gambl. Stud.* **13**, 253–266 (1997).
20. Wardle, H., John, A., Dymond, S. & McManus, S. Problem gambling and suicidality in England: secondary analysis of a representative cross-sectional survey. *Public Health* **184**, 11–16 (2020).
21. Karlsson, A. & Håkansson, A. Gambling disorder, increased mortality, suicidality, and associated comorbidity: a longitudinal nationwide register study. *J. Behav. Addict.* **7**, 1091–1099 (2018).
22. Wong, P. W. C., Chan, W. S. C., Conwell, Y., Conner, K. R. & Yip, P. S. F. A psychological autopsy study of pathological gamblers who died by suicide. *J. Affect. Disord.* **120**, 213–216 (2010).
23. LaPlante, D. A., Nelson, S. E., LaBrie, R. A. & Shaffer, H. J. Stability and progression of disordered gambling: lessons from longitudinal studies. *Can. J. Psychiatry* **53**, 52–60 (2008).
24. Wohl, M. J. & Sztainert, T. Where did all the pathological gamblers go? Gambling symptomatology and stage of change predict attrition in longitudinal research. *J. Gambl. Stud.* **27**, 155–169 (2011).
25. Browne, M., Goodwin, B. C. & Rockloff, M. J. Validation of the Short Gambling Harm Screen (SGHS): a tool for assessment of harms from gambling. *J. Gambl. Stud.* **34**, 499–512 (2018).
26. Browne, M. et al. *Assessing Gambling-related Harm in Victoria: A Public Health Perspective* (Victorian Responsible Gambling Foundation, 2016).
27. Canale, N., Vieno, A. & Griffiths, M. D. The extent and distribution of gambling-related harms and the prevention paradox in a British population survey. *J. Behav. Addict.* **5**, 204–212 (2016).
28. Salonen, A. H., Hellman, M., Latvala, T. & Castrén, S. Gambling participation, gambling habits, gambling-related harm, and opinions on gambling advertising in Finland in 2016. Nordisk Alkohol Nark. 35, 215–234 (2018).
29. Langham, E. et al. Understanding gambling related harm: a proposed definition, conceptual framework, and taxonomy of harms. *BMC Public Health* **16**, 80 (2016).
30. Wardle, H., Reith, G., Best, D., McDaid, D. & Platt, S. *Measuring Gambling-related Harms: A Framework for Action* (Gambling Commission, 2018).
31. Browne, M. & Rockloff, M. J. Prevalence of gambling-related harm provides evidence for the prevention paradox. *J. Behav. Addict.* **7**, 410–422 (2018).
32. Collins, P., Shaffer, H. J., Ladouceur, R., Blaszczynski, A. & Fong, D. Gambling research and industry funding. *J. Gambl. Stud.* **36**, 1–9 (2019).
33. Delfabbro, P. & King, D. L. Challenges in the conceptualisation and measurement of gambling-related harm. *J. Gambl. Stud.* **35**, 743–755 (2019).
34. Browne, M. et al. What is the harm? Applying a public health methodology to measure the impact of gambling problems and harm on quality of life. *J. Gambl. Issues* https://doi.org/10.4309/jgi.2017.36.2 (2017).
35. Markham, F., Young, M. & Doran, B. The relationship between player losses and gambling-related harm: evidence from nationally representative cross-sectional surveys in four countries. *Addiction* **111**, 320–330 (2016).
36. Markham, F., Young, M. & Doran, B. Gambling expenditure predicts harm: evidence from a venue-level study. *Addiction* **109**, 1509–1516 (2014).

37. Blaszczynski, A., Ladouceur, R. & Shaffer, H. J. A science-based framework for responsible gambling: the Reno model. *J. Gambl. Stud.* **20**, 301–317 (2004).
38. Rose, G. *The Strategy of Preventive Medicine* (Oxford Univ. Press, 1992).
39. *Improving the Financial Health of the Nation* (Financial Inclusion Commission, 2015); https://www.financialinclusioncommission.org.uk/pdfs/fic_report_2015.pdf

Acknowledgements

We thank A. Trendl and H. Wardle for comments on an earlier draft of this manuscript. We thank R. Burton, Z. Clarke, C. Henn, J. Marsden, M. Regan, C. Sharpe and M. Smolar from Public Health England and L. Balla, L. Cole, K. King, P. Rangeley, H. Rhodes, C. Rogers and D. Taylor from the Gambling Commission for providing feedback on a presentation of this work. We thank A. Akerkar, D. Collins, T. Davies, D. Eales, E. Fitzhugh, P. Jefferson, T. Bo Kim, M. King, A. Lazarou, M. Lien and G. Sanders for their assistance. We thank the Customer Vulnerability team, with whom we worked as part of their ongoing strategy to help vulnerable customers. We acknowledge funding from LBG, who also provided us with the data but had no other role in study design, analysis, decision to publish or preparation of the manuscript. The views and opinions expressed are those of the authors and do not necessarily reflect the views of LBG, its affiliates or its employees. We also acknowledge funding from Economic and Social Research Council (ESRC) grants nos. ES/P008976/1 and ES/N018192/1. The ESRC had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

P.P. and P.N. proposed the initial concept. All authors contributed to the design of the analysis and the interpretation of the results. J.G. and N.S. wrote the initial draft; all authors contributed to the revision. N.M. and P.P. constructed variables and N.M. prepared all figures and tables. D.L. established collaboration with LBG. D.L., J.G. and N.S. secured funding for the research. P.N. conducted a review of the existing literature.

Competing interests

N.M. was previously, and D.L. is currently, an employee of LBG. P.P. was previously a contractor at LBG. They do not, however, have any direct or indirect interest in revenues accrued from the gambling industry. P.N. was a special advisor to the House of Lords Select Committee Enquiry on the Social and Economic Impact of the Gambling Industry. In the last 3 years, P.N. has contributed to research projects funded by GambleAware, Gambling Research Australia, NSW Responsible Gambling Fund and the Victorian Responsible Gambling Foundation. In 2019, P.N. received travel and accommodation funding from the Spanish Federation of Rehabilitated Gamblers and in 2020 received an open access fee grant from Gambling Research Exchange Ontario. All other authors have no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-020-01045-w.

Correspondence and requests for materials should be addressed to N.M.

Peer review information Peer reviewer reports are available. Primary Handling Editor: Aisha Bradshaw.

nature research

Corresponding author(s): Naomi Muggleton

Last updated by author(s): Dec 7, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data were extracted from Lloyds Banking Group (LBG) databases using Teradata SQL Assistant (Version 15.10.1.9). The SQL code that supports the analysis reported here are available from LBG but restrictions apply to the availability of this code, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of LBG.
Data analysis	Data analysis was conducted using R (Version 3.4.4). The R code that supports this analysis can be found at github.com/nmuggleton/gambling_related_harm. Commercially sensitive code has been redacted. This should not affect the interpretability of the code.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from LBG but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available from the authors upon reasonable request and with permission of LBG.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	In this article we perform quantitative analysis on transactional data from a retail bank.
Research sample	The research samples consisted of LBG customers. Samples 1 and 3 consisted of a randomly selected, representative sample of 100,000 customers. Sample 2 consisted of all LBG customers who had an active personal current account and / or credit card in 2013 (~6.5 million individuals).
Sampling strategy	For Samples 1 and 3, we selected a representative, random sample of 100,000 active LBG customers. For Sample 2, we included all active customers.
Data collection	The data were provided by LBG.
Timing	Not applicable - the data were not collected by the authors.
Data exclusions	No data were excluded from the analysis.
Non-participation	Not applicable - we did not recruit participants, which means that no participants dropped out.
Randomization	Participants were not allocated to experimental groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging