Using causal models to distinguish between neurogenesis-dependent and -independent effects on behaviour

Stanley E. Lazic*
Bioinformatics and Exploratory Data Analysis, F. Hoffmann-La Roche
4070 Basel, Switzerland

There has been a substantial amount of research on the relationship between hippocampal neurogenesis and behaviour over the past 15 years, but the causal role that new neurons have on cognitive and affective behavioural tasks is still far from clear. This is partly due to the difficulty of manipulating levels of neurogenesis without inducing off-target effects, which might also influence behaviour. In addition, the analytical methods typically used do not directly test whether neurogenesis mediates the effect of an intervention on behaviour. Previous studies may have incorrectly attributed changes in behavioural performance to neurogenesis because the role of known (or unknown) neurogenesis-independent mechanisms was not formally taken into consideration during the analysis. Causal models can tease apart complex causal relationships and were used to demonstrate that the effect of exercise on pattern separation is via neurogenesis-independent mechanisms. Many studies in the neurogenesis literature would benefit from the use of statistical methods that can separate neurogenesis-dependent from neurogenesis-independent effects on behaviour.

Keywords: Bayesian graphical model; behaviour; causal model; hippocampus; neurogenesis; stem cells

1. INTRODUCTION

One of the advantages of laboratory-based studies is that it is often possible to experimentally control important variables that might influence the system under investigation. This is a luxury that some fields of biology (e.g. ecology) and many social sciences do not have. Instead, they often have to measure these relevant variables and use more complex statistical models to understand cause-and-effect relationships. Occasionally, some aspect of the system is not under complete experimental control, and it is necessary to use analytical methods developed in other fields to address key research questions. Such methods are often necessary when relating adult hippocampal neurogenesis to behaviour because factors that influence neurogenesis such as age, hormones, stress, physical activity, environmental enrichment, anti-depressants, disease status, etc. have off-target effects, and therefore establishing a causal relationship between neurogenesis and behaviour is not straightforward. Methods have been developed to test whether hypothesized causal relationships such as A affects B, and B affects C (but A does not affect C directly) are consistent with the data (denoted as $A \rightarrow B \rightarrow C$). Here, $B$ is said to mediate the effect of $A$ on $C$, and thus $A$’s effect on $C$ is indirect. However, a direct $A \rightarrow C$ effect might exist that does not involve $B$. How would one distinguish between these two situations? Or can $A$ have an effect on $C$ both via $B$ and directly? If so, what proportion of the effect is via $B$? Causal models can be used to address these types of questions, but are rarely employed in studies of adult hippocampal neurogenesis. This means that the analyses do not directly address the question that the study was designed to test, and incorrect conclusions can be reached regarding how $A$ affects $C$ (i.e. directly, via $B$, or both).

In a recent article, Creer et al. [1] examined the relationship between neurogenesis and performance on a pattern separation task. This task had two outcomes that were measured: the number of trials required to reach a set criterion, and the number of reversals achieved after a fixed number of attempts. Levels of neurogenesis were manipulated by placing a running wheel in the cages of some animals (runners), while controls had no running wheel. They showed that there was a relationship between neurogenesis and behaviour for the reversal data, and a non-significant trend in the predicted direction for the trials-to-criterion data. They concluded that because exercise (i) increased levels of neurogenesis, and (ii) led to better performance on the behavioural task, that neurogenesis might be important for this task. However, this conclusion does not follow from the analysis because the effect of exercise on other aspects of neural functioning have not been taken into consideration. In other words, they concluded that $A$ (exercise) affects $B$ (neurogenesis), which affects $C$ (behaviour). What about the direct

*stan.lazic@cantab.net

A → C effect, which subsumes all of the neurogenesis-independent effects? Could this completely account for the observed results?

Figure 1 shows two potential models of the causal relationships (indicated by arrows) between the treatment, neurogenesis and behaviour. In the first model, the treatment (exercise versus control) influences neurogenesis, which in turn influences behaviour (performance on a pattern separation task). In the second model, neurogenesis does not have a direct causal effect on behavioural performance. The treatment influences neurogenesis, and it influences behaviour directly (or more likely via some other variable that was not measured, such as changes in the electrophysiological properties of the cells, spine density, etc.). These two models make different predictions about relationships that will be found in the data, and therefore these predictions can be tested against the data. The model that correctly predicts the data will have the greater support, and as will be shown below, the evidence is strongly in favour of the neurogenesis-independent model. In addition to the two different predictions, these models also make three predictions which are in common, and therefore these three predictions cannot be used to discriminate between the models. Unfortunately, many published studies only test the common predictions, and therefore they cannot provide support for a causal role for neurogenesis.

2. METHODS

Data were accurately extracted from figures 2b and 3c of the original publication [1] using g3data software (www.frantz.fi/software/g3data.php). This re-analysis focuses mainly on the reversal data, since it had a significant relationship with neurogenesis. Analysis was conducted with R v. 2.12.2 [2,3]. In order to confirm the accuracy of the data extraction, the same analysis as Creer et al. was performed. \( R^2 \) between neurogenesis and behavioural performance was determined to be 0.235, which compares favourably with the results reported in the original publication (\( R^2 = 0.236; \) percent error less than 0.5%). The data and R code are provided in the electronic supplementary material, so that the analysis is documented and reproducible.

Two methods were used in the re-analysis of the data. The first is from Baron & Kenny [4], which can be implemented with a standard software, does not
require advanced statistical knowledge, and is even discussed in introductory statistics textbooks [5]. More recent developments and caveats of using this approach are discussed in references [6–14]. In this simple experiment (only three variables with linear relationships), the relevant questions can be addressed with t-tests, linear regression and an analysis of covariance (ANCOVA)—all standard techniques that will be familiar to many biologists. An ANCOVA model is characterized by a continuous response variable with a combination of continuous and categorical predictor variables. In this case, performance on the pattern separation task was the response variable, neurogenesis was the continuous predictor and treatment (exercise versus control) was the categorical factor. In the output of such an analysis, the effect of neurogenesis is adjusted for the effect of the treatment, and the effect of the treatment is adjusted for the effect of neurogenesis, and these results are conveniently the two predictions which differ between the models. These tests are all just specific examples of a linear model, and all of the above tests can be conducted within a linear modelling framework, which is a more unified method for data analysis [5,15,16]. All of the results are reported as regression coefficients from linear models, but the R code in the electronic supplementary material contains the equivalent results presented as t-tests, regression and ANCOVA. The key relationships are described with the following equations, where $T$, $N$ and $B$ represent the variables Treatment, Neurogenesis and Behaviour, respectively.

\[
N = \alpha_1 + \beta_1 T + \epsilon_1, \quad (2.1)
\]

\[
B = \alpha_2 + \beta_2 T + \epsilon_2, \quad (2.2)
\]

\[
B = \alpha_3 + \beta_3 N + \epsilon_3 \quad (2.3)
\]

and

\[
B = \alpha_4 + \beta_4 T + \beta_5 N + \epsilon_4. \quad (2.4)
\]

Figure 2. Neurogenesis and behavioural data. (a) Runners had significantly more new cells ($\beta_1, p < 0.001$), and (b) more reversals ($\beta_2, p = 0.002$) compared with the control group. Most papers in the literature that used a similar design stop the analysis here, incorrectly believing that this provides evidence that neurogenesis affects behaviour. There was also a significant correlation (c) between the two outcomes ($\beta_3$, solid line, $p = 0.030$), but this is still insufficient to indicate that neurogenesis affects behaviour. The dashed lines show the (lack of a) relationship between neurogenesis and behaviour conditioned on treatment group ($\beta_5$). Posterior densities from the Bayesian analysis (d) show that neurogenesis had no effect on the number of reversals conditioned on treatment group ($\beta_4$, dashed line, corresponds to dashed lines in panel c). Decomposing the total effect of exercise into the indirect effect (IE; via neurogenesis, $\beta_1 \times \beta_5$) and the direct effect (DE; neurogenesis-independent, $\beta_4$) shows that neurogenesis-independent mechanisms are affecting behaviour ($p = 0.016$). The peak of a distribution is the best estimate of the effect, and the width corresponds to the uncertainty in this estimate.
Equations (2.1)–(2.3) test model predictions 1–3 in figure 1, and equation (2.4) tests model predictions 4 and 5. The $\alpha$ parameters represent the intercept for each model and are not of direct interest. The $\beta$ parameters are the coefficients for the effect in question and the subscript corresponds to the model prediction that is being tested. The $\epsilon$’s are the error terms (residuals). In the context of this experiment, the main interest is in the relationship between neurogenesis and behaviour, once the effect of exercise on behaviour has been removed (i.e. testing whether $\beta_e$ is different from zero). A related question is whether the effect of exercise is mediated via neurogenesis (for example, if the main research question is ‘By what mechanism does exercise affect behaviour?’). This can be determined by requiring that (i) the treatment affects neurogenesis and (ii) neurogenesis is associated with behaviour after adjusting for a direct effect (DE) of treatment on behaviour (i.e. $\beta_i$ and $\beta_e$ both need to be significant). However, this piecemeal approach is not ideal because two statistical tests are performed, and the strength of the relationship is not quantified. An alternative is the product-of-coefficients method, where $\beta_i \times \beta_e$ needs to be significantly different from zero. One difficulty with this approach is estimating the standard error, or uncertainty, of the estimate [6,7]; however, this can be easily estimated in the Bayesian analysis below.

The second method uses a Bayesian graphical model to estimate causal relationships. This is a general method, which can be scaled up to more complex experimental designs, provides more robust estimates with small samples sizes, and allows latent variables and prior information from previous experiments to be included. In addition, direct probability statements can be made about the parameters, and the results can provide support for the null hypothesis [17]; traditional null hypothesis significance testing can only reject (or fail to reject) the null hypothesis, but cannot directly provide support for it. Graphical models are an active area of research with developments occurring in many fields, and as often happens, similar methods are developed with different names (e.g. structural equation modelling, path analysis, Bayesian networks, probabilistic graphical models and causal mediation analysis), along with slightly different philosophies, emphases, assumptions and underlying algorithms [18–24]. A general introduction to probabilistic graphical models using gene expression and cell signalling as examples is provided by Needham et al. [25,26] and Friedman [27], and a more detailed description of these methods using a simple three-variable problem can be found in Yuan & MacKinnon [23]. Briefly, the hypothesized relationships are written as directed acyclic graphs (DAGs; figure 1), where each variable is represented as a node, and hypothesized causal relationships are represented by arrows. These graphical representations are then converted into a set of equations (similar to equations (2.1) and (2.4)).

The joint distribution of all of the variables is factored into a set of simpler conditional distributions, much like the number 12 can be factored into $3 \times 2 \times 2$. Graphical models with a different arrangement of hypothesized causal relationships will be factored in different ways (e.g. 12 can also be factored into $6 \times 2$ and $4 \times 3$), following a few simple rules. If a variable has no arrows pointing into it (i.e. it is not dependent on another variable), then it is written as a regular probability density or distribution function (e.g. $P(X)$ is the probability density function of a continuous variable $X$). Variables that have an arrow pointing into them are condition on those variables. For example, if there is an arrow pointing from $Y$ to $X$, this would be written as $P(X|Y)$, which is read as the probability of $X$ given $Y$ (the ‘$|$’ means ‘given’). Thus, the joint probability of the three variables $P(T,N,B)$ can be rewritten as a set of conditional distributions, which correspond to the two graphical models (figure 1). For model 1, this can be written as

$$P(T,N,B) = P(B|N)P(N|T)P(T),$$  \hspace{1cm} (2.5)

whereas model 2 would be written as

$$P(T,N,B) = P(B|T)P(N|T)P(T),$$  \hspace{1cm} (2.6)

with the only difference being whether $B$ is conditioned on $N$ or $T$, reflecting the two potential variables that affect behaviour. It is then possible to test whether there is any evidence in the data for the hypothesized causal structures represented by these equations, and to quantify the degree to which one model is better compared with the other.

Analysis was conducted with the R2WinBUGS R package [28] and OpenBUGS, and the code is provided in the electronic supplementary material. Non-informative priors were used and the results were not sensitive to the form of the prior (e.g. normal versus uniform). Three chains with 500,000 iterations each were used for the Markov chain Monte Carlo (MCMC) sampling, with a burn-in period of 250,000 iterations, and every tenth value was saved. The three chains were well mixed (Gelman–Rubin statistic less than 1.01 for all parameters). The main parameter of interest was the effect of neurogenesis on behaviour after adjusting for the indirect effect (IE) of exercise on behaviour ($\beta_i$). In addition, the direct ($\beta_i$) and indirect ($\beta_i \times \beta_e$) effects of exercise on behaviour were determined.

Many of the requirements for using causal models are the same as for a standard analysis, including a large enough sample size to obtain reasonably precise estimates, important variables have not been omitted, random assignment of animals to groups, and randomized collection of data, processing of samples and quantification to avoid time-dependent confounders. Causal models have a few additional assumptions. First, the mediator (i.e. neurogenesis) should ideally be measured without error [4]. Creer et al. expressed the neurogenesis data as a density measurement (neurons mm$^{-3}$), and so the extent to which values represent changes in cell number (which is of interest) or changes in the volume of the dentate gyrus is not clear. It is assumed that the reported estimates are an accurate representation of the number of cells and are not influenced by changes in volume between conditions. However, such assumptions can be misleading and are unnecessary when using modern design-based stereological methods [29–31]. Second, there should be no neurogenesis by treatment interaction, which was
reasonable to assume with these data \(F_{1,16} = 0.17, p = 0.687\), although more recent methods can handle interactions [13]. Third, the residuals \(\epsilon_1\) and \(\epsilon_4\) should not be correlated, and levels of neurogenesis and \(\epsilon_4\) should not be correlated, as this might indicate that another variable not included in the analysis is affecting levels of neurogenesis and behaviour (both correlations were almost zero \(p > 0.999\)). Finally, it is assumed that the arrows are pointing in the right direction. This may be difficult to check in general, but because this was a randomized experiment, we can be certain that levels of neurogenesis and behavioural performance did not affect assignment to different treatment groups.

3. RESULTS

3.1. Predictions common to both models

The first prediction common to both models is that levels of neurogenesis will differ between the runners and control mice. This is indeed the case, with mice in the exercise condition having approximately twice as many new cells (figure 2a, \(\beta_1 = 3.23, 95\%\ CI = 2.03\) to 4.43, \(p < 0.001\)).

The second prediction common to both models is that performance on the behavioural task will differ between the runners and control mice. The neurogenesis-dependent model predicts this because the treatment influences neurogenesis, and levels of neurogenesis in turn influence performance on the pattern separation task. The neurogenesis-independent model predicts this because exercise influences pattern separation directly (or more likely via some unmeasured variable). The runners had more than twice as many reversals compared with the control group (figure 2b, \(\beta_2 = 0.96, 95\%\ CI = 0.40\) to 1.52, \(p = 0.002\)), which is consistent with both models.

The third prediction common to both models is that there will be a relationship between neurogenesis and behaviour. The neurogenesis-dependent model predicts this because neurogenesis has a causal influence on pattern separation. This is the main research question that the study was designed to test. The problem is that the neurogenesis-independent model makes the exact same prediction. This is because neurogenesis and behaviour are influenced by the same variable (the treatment) that induces a correlation between them. This is the classical third-variable problem, where A and B are correlated, but only because they share a common cause in variable C. The relationship between neurogenesis and behaviour is significant (slope of the solid line in figure 2b, \(\beta_3 = 0.18, 95\%\ CI = 0.02\) to 0.34, \(p = 0.030\)), and Creer et al. took this as evidence that neurogenesis might be involved in the pattern separation task. However, since the neurogenesis-independent model makes the same prediction, further analyses are required to separate the neurogenesis-dependent from the neurogenesis-independent effects.

3.2. Predictions that discriminate between models

If neurogenesis is causally involved in the pattern separation task, then blocking the increase in neurogenesis caused by exercise will cause the performance of the runners and controls to be the same (that is, the relationship between treatment and behaviour will no longer be significant). We could imagine doing this experimentally, where the mice would be given a compound whose sole effect is to nullify the effect of exercise on neurogenesis (indeed, this approach is often taken, and studies with this design are discussed below). Exercise would not affect behaviour in this case because the causal link between treatment and behaviour has been broken by holding neurogenesis at a constant level. The neurogenesis-independent model, however, predicts that the relationship between treatment and behaviour will still exist after keeping the levels of neurogenesis constant. This is clear from figure 1; in the neurogenesis-independent model, there is a direct causal link between the treatment and behaviour, so whether levels of neurogenesis are held fixed or not is irrelevant. It is also possible to statistically hold constant, or fix, levels of neurogenesis at some value (also referred to as ‘adjusting for’, ‘taking into account’, ‘controlling for’ or ‘conditioning on’), and when this is done, we find that there is still a significant relationship between treatment and behaviour (difference between runners and controls: \(\beta_4 = 1.06\) reversals, 95\% CI = 0.09 to 2.02, \(p = 0.034\)), thus providing support for the neurogenesis-independent model.

The other prediction that differs between the models (prediction five in figure 1) is that if the neurogenesis-independent model is correct, then conditioning on the treatment will remove the correlation between neurogenesis and behaviour. This is clear from figure 1, where it can be seen that in the neurogenesis-independent model, the link between neurogenesis and behaviour will be broken if we condition on the treatment. This can be thought of as removing a common cause of neurogenesis and behaviour, and testing whether there is still a relationship between them. However, the neurogenesis-dependent model predicts that there will still be a relationship between neurogenesis and behaviour because there is a direct causal link between the two. The results show that there is no relationship between neurogenesis and behaviour when conditioned on treatment (\(\beta_5 = -0.03, 95\%\ CI = -0.27\) to 0.21, \(p = 0.788\), represented as the slope of the two dashed lines in figure 2c), which also supports the neurogenesis-independent model. Note that this is not owing to a lack of statistical power, the value is close to zero, and is even in the opposite direction to what the neurogenesis-dependent model predicts (i.e. higher levels of neurogenesis were associated with worse performance, although not significantly so). This model is actually fitting a separate regression line for the runners and controls, rather than one regression line through all of the data points as in figure 2c. The relationship between neurogenesis and behaviour needs to hold within each group, and should not be driven by differences between the group means of the two variables.

The above analysis examined individual parameters to see whether they were different from zero, which is useful to address particular questions about the data. It is also possible to examine a model as a whole, by calculating a measure of model fit or adequacy. Two...
models can be fit to the same data and then compared, and the model with the better fit is preferred. Models that are more complex (have more parameters) have greater freedom to better approximate the data and therefore will fit better than less complex models. Therefore, the complexity of the models must also be taken into account when performing a comparison (in this case the two models had the same number of parameters). A number of methods have been developed for this type of model comparison, and the deviance information criterion (DIC; [32]) is one appropriate option for the Bayesian analysis. The lower the DIC for a model, the better the fit, and the larger the difference in DIC’s between two models, the better one model is versus the other. There are no strict rules regarding how large a difference is considered important, however a difference in DIC between 5 and 10 can be considered substantial (http://www.mrc-bsu.cam.ac.uk/bugs/winbugs/dicpage.shtml). The DIC for the neurogenesis-dependent model was 116.3, while for the neurogenesis-independent model was 110.9. The difference of 5.4, therefore, provides strong support for the neurogenesis-independent model.

The above analysis pitted two competing models against each other that hypothesized different mechanisms for the effect of exercise on behaviour. While this is useful for explanatory purposes, it is likely that multiple mechanisms are at work (both neurogenesis and neurogenesis-independent), and the main research question should be ‘to what extent does neurogenesis contribute to behaviour’? In other words, it would be useful to determine how much of exercise’s effect is due to neurogenesis, and how much is due to neurogenesis-independent mechanisms. The relationship between exercise and behaviour was, therefore, decomposed into the indirect/neurogenesis-dependent (Exercise → Neurogenesis → Behaviour) and direct/neurogenesis-independent (Exercise → Behaviour) paths, which were then estimated. This can be represented as

\[
P(T, N, B) = P(B | T, N)P(N | T)P(T),
\]

where the first term now has B conditioned on both T and N (compare with equations (2.5) and (2.6)). The results are displayed in figure 2d, which show the posterior densities for the two effects. These density plots represent the estimated effects along with the uncertainty associated with them. For example, the DE of exercise via neurogenesis-independent mechanisms is to increase the number of reversals by approximately one (the peak of the distribution), with a 95% credible interval of 0.09–2.03 reversals (a 95% credible interval means that there is a 95% chance that the true effect is between the upper and lower values, this is often what confidence intervals are mistakenly believed to be). Since the lower interval excludes zero, we can conclude that there are neurogenesis-independent mechanisms at work (the associated p-value is 0.016). However, the IE of exercise via neurogenesis on behaviour was very close to zero, and was slightly negative. In other words, the ‘best guess’ of the effect of increasing neurogenesis is that it worsens performance (which is the same result as the first analysis). If the effect of neurogenesis is actually zero, a negative value would be expected 50 per cent of the time, so this is not surprising. It is clear from these results that exercise is not affecting behaviour via changes in levels of neurogenesis.

The results are the same for the trials-to-criterion data. Creer et al. established that the first two predictions are supported by the data. The third prediction (a relationship between neurogenesis and behaviour) was not significant (\( \beta_3 = -3.79, 95\% \text{ CI} = -8.74 \text{ to } 1.15, p = 0.125 \)), but they described it as a ‘trend’. However, this relationship between neurogenesis and behaviour is greatly attenuated (\( \beta_3 = -0.16, 95\% \text{ CI} = -8.37 \text{ to } 8.04, p = 0.967 \)) when conditioned on treatment, indicating that there is little support for the neurogenesis-dependent model.

Creer et al. concluded that ‘Taken together, our findings indicate that exercise-induced neurogenesis improves dentate gyrus-mediated encoding of distinct spatial representations’, and it seems that this conclusion has been accepted by others [33–37]. However, a re-analysis of their data using causal models—which can tease apart complex relationships and directly test whether neurogenesis mediates the effects of exercise on behaviour—clearly shows that the effect of exercise is via neurogenesis-independent mechanism(s). This is further supported by their finding in a separate experiment, where aged runners (the reanalysed experiment used younger adults) did not have an increase in neurogenesis, but still showed a small improvement on a modified version of the pattern separation task, which is exactly what the neurogenesis-independent model predicts. It should also be noted that the adult mice ran \( 23.5 \pm 1.79 \text{ km d}^{-1} \), while the aged mice only ran \( 5.4 \pm 0.68 \text{ km d}^{-1} \). Therefore, because aged animals ran 77 per cent less, the effect of running (via neurogenesis-independent mechanisms) will be predicted to influence behaviour to a lesser extent. While this manuscript was under review, this group has published a review article discussing other factors that may mediate the effects of exercise and environmental enrichment on behaviour [38]. We need to go further however, and convert these pictorial representations of hypothesized causal relationships into statistical models which can be tested against the data.

4. DISCUSSION

Causal models are rarely used to analyse data from laboratory-based biological experiments, which is partly due to the high experimental control that can usually be achieved in such situations (lack of knowledge of these methods is another reason). However, it is not always possible to create two experimental groups that differ only on the variable of interest. For example, the runners are not the same as the controls in all respects (even though they may have been at baseline owing to randomization), with the only exception of having higher levels of neurogenesis. If these other variables are potential candidates for mediating the effect of the treatment, then it is not possible to make causal statements about the variable of interest. One could imagine the exact same study (and analysis) carried out by a second group whose interest is not in

\[ \]
neurogenesis, but in spine density (also affected by exercise [39,40]), which was measured instead. Another group would look at levels of synaptic proteins, while a fourth would focus on glutamate receptors (both of which are also affected by exercise [41,42]). Without the appropriate analysis, each group would conclude that their hypothesized variable is important for pattern separation, and all four would only have an incomplete picture. It is possible that a particular variable has no effect, and thus the wrong conclusion would be reached. If more than one variable is important, then all of the estimated effects will be biased because the total effect (via all mechanisms) is being estimated, rather than the individual effects separately.

While there was no evidence for a causal role of neurogenesis in the present experiment, this does not rule out a contribution of neurogenesis to other behavioural tasks, or for the effect of decreasing neurogenesis on pattern separation [43]. Indeed, the causal relationships are likely to be more complex than suggested by this analysis, since angiogenesis was also affected by the treatment, and this introduces another variable by which the treatment might affect behaviour [44].

4.1. Manipulating neurogenesis often affects other relevant variables

If the goal of a study is only to demonstrate the effect of exercise on behaviour, then the concerns discussed herein do not apply. However, problems arise if the interest is in the effect of neurogenesis on behaviour, and exercise (or some other treatment) is used as a means to manipulate neurogenesis. Most methods of manipulating neurogenesis also affect other variables that are known (or could reasonably be expected) to contribute to behaviour (table 1).

For example, exercise affects spine density in the entorhinal cortex, CA1 and dentate gyrus [39,40], as well as levels of synaptic proteins [41], glutamate receptors [42] and various other genes, growth factors and neurotransmitters in the dentate gyrus [47,48]. If other factors are known to be involved, it is not possible to address the question ‘what is the effect of neurogenesis on behaviour?’ simply by looking at the relationship between the two. This will lead to severely biased results if the other relevant factors are not taken into account, as was demonstrated in the present study. Once it is known that exercise affects other aspects of the brain that might provide alternative causal explanations, future studies must take them into account, or at least rule them out as a potential contributing factor in any particular study. In addition to experimental manipulation, some studies use groups that naturally vary in levels of neurogenesis, such as young versus old, or a disease model versus control. Here, neurogenesis is not manipulated directly, but clearly there are many differences between old and young brains, or in transgenic/knock-out mice versus controls that need to be ruled out. Importantly, anything that affects physical activity or locomotor behaviour in the home cage (e.g. a disease model, general health and age) might then influence performance on a behavioural task.

Table 1. A selection of off-target effects for commonly used methods of manipulating neurogenesis. Owing to these off-target effects, it is not possible to make causal claims regarding the role that neurogenesis has on behaviour by simply examining the relationship between neurogenesis and behaviour—more complex models are needed.

<table>
<thead>
<tr>
<th>Off-target effects</th>
<th>references</th>
</tr>
</thead>
<tbody>
<tr>
<td>stress/corticosterone</td>
<td>hippocampal volume [45]</td>
</tr>
<tr>
<td>environmental enrichment</td>
<td>dendrites/spines [49–51]</td>
</tr>
<tr>
<td>MAM*</td>
<td>general health [53]</td>
</tr>
<tr>
<td>imipramine</td>
<td>dendrites/synapses [45,54,55]</td>
</tr>
<tr>
<td>fluoxetine</td>
<td>dendrites/spines [45,56]</td>
</tr>
<tr>
<td>irradiation</td>
<td>NMDA receptors [57]</td>
</tr>
</tbody>
</table>

*Methylazoxymethanol acetate.

This re-analysis not only reverses the conclusions of the original study, but also brings into question the conclusion of many published studies purporting to demonstrate evidence of a causal association between neurogenesis and behaviour. This is because the basic design, analysis and logic of the Creer et al. study is similar to many studies in the literature: manipulate neurogenesis and observe behaviour; if there are differences in both neurogenesis and behaviour between groups, conclude (tentatively) that neurogenesis might be causally involved. Furthermore, ignore other potential variables that might explain the results. The details of the studies vary (e.g. method of manipulating neurogenesis, type of behavioural task, etc.), but they follow the same basic template. The onus is on those reporting associations between neurogenesis and behaviour to demonstrate that neurogenesis-independent mechanisms were not at work, or at least quantify their magnitude and estimate the unique contribution that neurogenesis makes. For many studies, this has not been done, and there is good reason to believe that neurogenesis-independent effects play a role in these studies (see table 1 for a brief list). It is possible (and indeed not difficult) to design studies with biased effects that can be consistently reproduced [64], and it is likely that this has occurred in the neurogenesis literature. The possibility that neurogenesis-independent effects might play a role is acknowledged in many studies and review articles (e.g. Creer et al. noted that exercise is known to affect expression of neurotrophins, vascularization, dendritic spine density and synaptic plasticity), but the functional consequences of this when analysing the
data and interpreting the results are generally ignored. If these off-target effects are not under experimental control, then the only option is to use more complex statistical models to account for them.

Many studies only establish that the first two predictions in figure 1 hold, and some studies such as Creer et al. test the third prediction, but given that these results are also entirely consistent with a system in which the behavioural task is completely independent of neurogenesis, these results cannot support the hypothesis that hippocampal neurogenesis influences behaviour. Therefore, statements such as ‘Thus, the positive correlation between enhanced neurogenesis and improved learning... remains a valid basis for the suggestion that newly generated cells may be important for memory function.’ [65 p. 129], and ‘Clearly, the results of the past 15 years of research support the idea of a functional role of adult-generated neurons in learning and memory processes...’[66 p. 385], need to be reconsidered, as they have less empirical support than commonly supposed. It should be noted that these are not poorly designed or flawed studies; the analysis simply does not directly test the main research question, or take into account potential confounding variables, which can lead to incorrect conclusions. Recent review articles comment on the conflicting results in the literature [67–71], and often suggest that this is due to some interaction between the method of manipulating neurogenesis, the method of measuring neurogenesis, the species or strain of animal used, differing demands of the behavioural task, the type of memory examined, etc. While these factors most probably play a role, conflicting results are not unexpected if important factors, often unmeasured, are influencing the results. It would be useful to reanalyse many of the key studies using causal models. These studies only estimated the total effect of their respective treatments on behaviour, and if any part of these effects was via neurogenesis-independent mechanisms, then the estimates were biased and attributed too much of the total effect to neurogenesis. If neurogenesis-independent effects were present, then separating the total effect into a direct (neurogenesis-independent) and indirect (neurogenesis-dependent) component would decrease the estimated influence of neurogenesis. Thus, studies that were initially statistically significant will have attenuated estimates and larger p-values. One can only speculate on the number of studies whose conclusions will have to be revised. To get a feeling for what this number might be, it is interesting to note that only five studies (including this re-analysis) [44,72–74] used models that conditioned on important variables, and these have either found no association between neurogenesis and behaviour, or found that higher levels of neurogenesis were associated with worse behavioural performance.

Recently, Sahay et al. [75] published a very thorough paper which used genetic methods to selectively increase the survival of new born neurons. They found that animals in the high neurogenesis group were better able to discriminate between two similar contexts, suggesting improved pattern separation. However, perhaps the most interesting result was that there were no differences between the normal and high neurogenesis groups on the standard behavioural tests that numerous previous studies showed to be sensitive to levels of neurogenesis, including the Morris water maze, novel object recognition, active place avoidance, forced swim test and novelty-suppressed feeding. This suggests that previous studies might have been incorrectly attributing differences in behavioural performance to neurogenesis, while the actual effects were via neurogenesis-independent mechanisms. It is likely that genetic methods of manipulating hippocampal neurogenesis are ‘cleaner’ in that there are fewer off-target effects, however this is an assumption, and one which can be tested with the data at hand. In other words, whether neurogenesis-independent effects exist, and the magnitude of these effects is an empirical question that can be easily addressed.

4.2. More complex designs

It is recognized that neurogenesis-independent factors can play a role, and some studies explicitly state that they used two methods of decreasing neurogenesis to avoid off-target effects [76]. In addition, some studies [77–83] go a step further and try to demonstrate causality by (i) increasing neurogenesis and showing this improves performance on a behavioural task, and then (ii) inhibiting this increase in neurogenesis and showing that the behavioural improvement is lost. This is doing experimentally what the causal models do statistically, and can provide stronger evidence for a causal role of neurogenesis. These hypothesized causal relationships are shown with solid arrows in figure 3. However, it is still possible for neurogenesis-independent mechanisms to completely explain such
results. This could occur if the method of increasing neurogenesis directly affects behaviour (e.g. Exercise → Behaviour dashed line in figure 3), and the method of decreasing neurogenesis also directly affects behaviour (e.g. Corticosterone → Behaviour dashed line), without neurogenesis playing a role (no Neurogenesis → Behaviour arrow in figure 3). For example, it is possible that exercise and corticosterone affect behaviour directly, and also affect neurogenesis as a by-product. The relationships that actually exist in the data can be tested, and there is no need to speculate about which model or interpretation of the data are correct. The use of causal models in such a situation provides a test of the assumption that neurogenesis-independent mechanisms are not operating, and if they are, their magnitude can be quantified. It should also be stressed that this requires no further experimentation, these relationships can be tested with the available data. The electronic supplementary material contains the results of a simulation study where it is demonstrated that using the standard analytical methods, it is not possible to distinguish between a situation where the behavioural outcome is completely dependent on neurogenesis from a situation where neurogenesis plays no role. The causal modelling approach can clearly distinguish these two situations. In addition, these graphical representations are intuitive, make the hypothesized relationships explicit, highlight the implied assumptions (e.g. independence assumptions between variables), and can be converted into statistical models that can be used to make inferences. Furthermore, such model building is the way science advances [84].

5. CONCLUSIONS

Nakagawa & Hauber [85] suggested five statistical methods that should be used more often by neuroscientists (meta-analysis, mixed-effects modelling, multiple imputation, model averaging and MCMC), and to this list we could add a sixth ‘M’: mediation analysis. One reason for the extensive research on hippocampal neurogenesis is that it is widely believed to play a role in some cognitive and affective behaviours. If this is true, then manipulating levels of neurogenesis is a logical approach for improving memory and treating depression. However, if neurogenesis does not have a causal role on behaviour, or if the role of neurogenesis is small compared with the neurogenesis-independent mechanisms of a treatment, then resources would be better spent elsewhere. It is necessary to establish whether observed relationships are causal, as well as their magnitude, especially if the ultimate aim is to manipulate the system for therapeutic ends. Simple methods exist and are routinely used in other fields, and there is no reason for their continued omission in the neurogenesis literature.

The comments and suggestions from four anonymous reviewers are gratefully acknowledged.

REFERENCES


19 Grace, J. B. 2006 Structural equation modeling and natural systems. Cambridge, UK: Cambridge University Press. (doi:10.1017/CBO9780511617799)


