

Research Article

Open Access

Paul H. Walton*, Daniel J. Walton

Simpson's Paradox in the interpretation of "leaky pipeline" data

DOI 10.1515/ijtr-2016-0013

received October 2015; accepted November 2015

Abstract: The traditional 'leaky pipeline' plots are widely used to inform gender equality policy and practice. Herein, we demonstrate how a statistical phenomenon known as Simpson's paradox can obscure trends in gender 'leaky pipeline' plots. Our approach has been to use Excel spreadsheets to generate hypothetical 'leaky pipeline' plots of gender inequality within an organisation. The principal factors, which make up these hypothetical plots, can be input into the model so that a range of potential situations can be modelled. How the individual principal factors are then reflected in 'leaky pipeline' plots is shown. We find that the effect of Simpson's paradox on leaky pipeline plots can be simply and clearly illustrated with the use of hypothetical modelling and our study augments the findings in other statistical reports of Simpson's paradox in clinical trial data and in gender inequality data. The findings in this paper, however, are presented in a way, which makes the paradox accessible to a wide range of people.

Keywords: Leaky pipeline; Simpson's paradox; Gender equality; Gender inequality; Hypothetical modeling

1 Introduction

A cornerstone of gender equality studies is the careful collection and accurate interpretation of statistical data. Furthermore, the development of effective equality policies demands an evidence base, which is, more often than not, founded on quantitative data. In this regard, "leaky pipeline" plots are very widely used as a means of indicating the extent of gender inequality across a broad range of scenarios. Indeed, such plots have been used almost exclusively to determine the extent of gender inequality within a particular sphere of study, be it within an organisation (Verniers et al., 2015; McAllister et al., 2016), a profession (Jensen et al., 2015), a nation (Conroy et al., 2015) or even over a continent (Shaik and Fusulier, 2015). The plots are easy to understand, compelling in their message and widely accepted as a principal indicator of the presence or absence of gender inequality. Since they are used to inform a myriad of management decisions it is essential that the shortcomings in these plots are widely known. Unfortunately, and of importance to gender equality practice, it is not true that leaky pipeline plots are reliable bases for the evaluation of gender inequality. Like nearly all other quantitative estimators of complex situations, these plots are subject to multiple confounding factors (Anderson et al., 2009; Pearl, 2009), the most surprising of which is a statistical phenomenon known as Simpson's paradox or Simpson's reversal. Simpson's paradox describes a situation across a wide range of statistical analyses, including leaky pipeline plots, in which data can often be interpreted in a way which is opposite to the actual situation on which the plots are reporting (Pearl, 2014). The dangers of falling foul of the paradox are self-evident for policymakers and equality managers.

Simpson's Paradox (Simpson, 1951), originally reported by Yule at the start of the last century (Yule, 1903), reveals that when individual component variables (continuous or categorical) are amalgamated into a single variable, the overall trend in the single variable may well be opposite to the individual trends of the component variables. This situation was brought to prominence in a famous study of apparent gender inequality in Berkeley

*Corresponding author: Paul Walton, Department of Chemistry, University of York, Heslington, York, YO10 5DD, UK. Email: paul.walton@york.ac.uk

Daniel J. Walton, Department of Economics, University of Bath, Claverton Down, Bath, BA2 7AY, UK

graduate school admissions (Bickel et al., 1975) in which it was demonstrated that individual academic departments did not necessarily have gender bias in their graduate admission programs, despite very clear differences in the overall percentage of female students who were admitted compared to male students. In other words, within the realm of equality, leaky pipeline plots have the potential to mislead equality practitioners and managers to such an extent that they may report that there is a serious gender inequality problem within an organisation when this is not the case, or—worse still—the opposite. Additionally, the highly counterintuitive nature of the paradox means that, even when people are aware of its existence, it is rejected as rational possibility (van der Lee and Ellemers, 2015; Pearl, 2009). A recent high profile example attests to exactly this occurrence (Volker and Steenbeek, 2015; Albers, 2015), in which an in-depth study on grant success rates by the Danish Research Funding Council found “compelling evidence” that funding rates were gender biased, a finding which was backed up by a battery of statistical significance tests. Despite this finding and the subsequent publication of the work in a highly esteemed scientific journal, it was quickly pointed out by others that the findings based largely on leaky pipeline data may have been an example of Simpson’s paradox and could not be used to disprove the null hypothesis. Accordingly, the investigators could not have categorically concluded that gender inequality existed, even though there may have been good cause to suspect that gender inequality was present during the grant allocation process. This particular example vividly demonstrates that unawareness of Simpson’s paradox for policy intervention, finances and management practice is very significant indeed. It is, therefore, of some importance that awareness of the paradox (and for that matter other potential paradoxes) is raised amongst equality practitioners and managers. This awareness-raising further needs to be coupled to a clear explanation of how and where the paradox occurs. In this regard, there have been significant previous attempts to highlight the importance of Simpson’s paradox in data interpretation, most notably in data associated with clinical trials (Norton and Divine, 2015; Fenton et al., 2015). It is a surprise, therefore, that the paradox continues to be something that is overlooked not only in clinical trials but also in gender equality data.

It is in this context that we present here a straightforward analysis of Simpson’s paradox described in terms of a hypothetical situation, in which we build-up the picture of a leaky pipeline plot from its individual contributing components. By doing so we lay bare some of the factors, which can lead to Simpson’s paradox being observed in

data. Our objective is to augment the existing commentaries of Simpson’s paradox in data interpretation by exemplifying its occurrence in a straightforward and intelligible example, which, in turn, can be easily related to situations faced by managers and equality practitioners. Our aim is not only to bring knowledge of the paradox more into mainstream gender equality thinking, but also to equip practitioners and managers with the mental tools to challenge critically their initial reaction to leaky pipeline plots and to ‘look beyond the data’ such that they can more accurately evaluate where gender inequality may or may not exist.

2 Discussion

In terms of gender equality work, leaky pipeline data are presented in the form of categorical variables in which each variable is presented as the percentage of female-associated participants in that particular category. Many examples of the use of ‘leaky pipeline’ plots to evaluate potential gender bias can be found, ranging from job applications and appointments through to promotions in an organisation and research grant funding rates. For instance the percentage of female staff within a particular grade at a university (research assistant, assistant professor, associate professor, full professor) could be used to show levels of potential inequality in progression through those grades (see Figure 1 for example). Evaluators of such plots are instinctively drawn to differences in percentages of women in various categorical variables as an obvious and clear sign of gender bias (Pearl, 2014). Indeed, more often than not, these plots show a decline in the percentage of women as the categorical variable progresses and this observation is very frequently used by managers and policymakers to conclude that gender bias exists. In fact, it is indeed possible that these plots most probably do reflect levels of gender inequality, given the very extensive evidence from multiple other studies which demonstrate gender inequality across recognition, progression and promotion of men and women in organisations (Valian, 1999; Moss-Racusin et al., 2012; West et al., 2013). Indeed, our intention is not to undermine these wider studies, which repeatedly show that differential barriers to the progression of men and women are a major factor in the decline in the numbers of women at the higher grades of any institution and/or organisation. The central aspect to our analysis is that leaky pipeline plots should not be used, by themselves, to support an unequivocal conclusion of gender inequality. We illustrate our approach by

constructing three different hypothetical examples. Our method has been to use a simple model to calculate the traditional 'leaky pipeline' plot (percentage of female staff as a fraction of the staff total at each separate grade) for a hypothetical university, similar to the normal data that a university senior management team might expect to receive. The principal variables in our model are the size of the individual university departments (n), the fraction of staff in one grade compared to the immediately previous grade, progression index (p), the percentage of female staff who 'begin' in a department on the most junior grade ($\%F$) and an inequality index (i) for each department. We define the inequality index as the difference in the percentage of men and women who progress from one grade to the next. For example, if 40% of men at grade 1 are promoted to grade 2, and 35% of women are promoted from grade 1 to grade 2, then the inequality index is 5%. These data are used to generate hypothetical leaky pipeline plots, which have a similar form to the leaky pipeline plots based on real data (Figure 1). The most notable difference between our models and real data is that the progression index from one grade to the next varies significantly in the real data depending on the grades (varies between 0.3 and 0.7) whereas we have assumed a fixed index for all grades.

2.1 Hypothetical model

Imagine a University of Utopia in which there are equal numbers of male and female academic staff (represented by n at 50% in Table 1), and that the promotion of men and women occurs at an equal rate across the full scale

of progression points from the research assistant (grade 1) to full professor (grade 4). This equality in progression is modelled using an inequality index of 0% (i , shown in Table 1). At the University of Utopia, the inequality index is 0%. Also at the University of Utopia the progression prospects from one grade to the next are good since, on average, 60% of staff at one grade are promoted to the next (represented by the progression parameter of 0.6, p , in Table 1). This situation is depicted simply in Figure 2a. This traditional 'leaky pipeline' plot (which is not leaky in this case) shows that the percentage of female staff at each career grade is 50%, in other words, all appears to be fair. A bubble plot (Figure 2b) further shows the numbers of women at each grade, again depicting the situation to be fair.

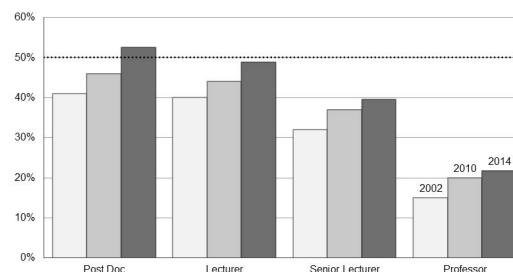


Figure 1: Traditional 'leaky pipeline' or scissors plot for UK universities, 2002 – 2014. Data from Equality Challenge Unit: *Equality in higher education: statistical report 2014*. http://www.ecu.ac.uk/wp-content/uploads/2014/11/2014-08-ECU_HE-stats-report_staff_v19.pdf

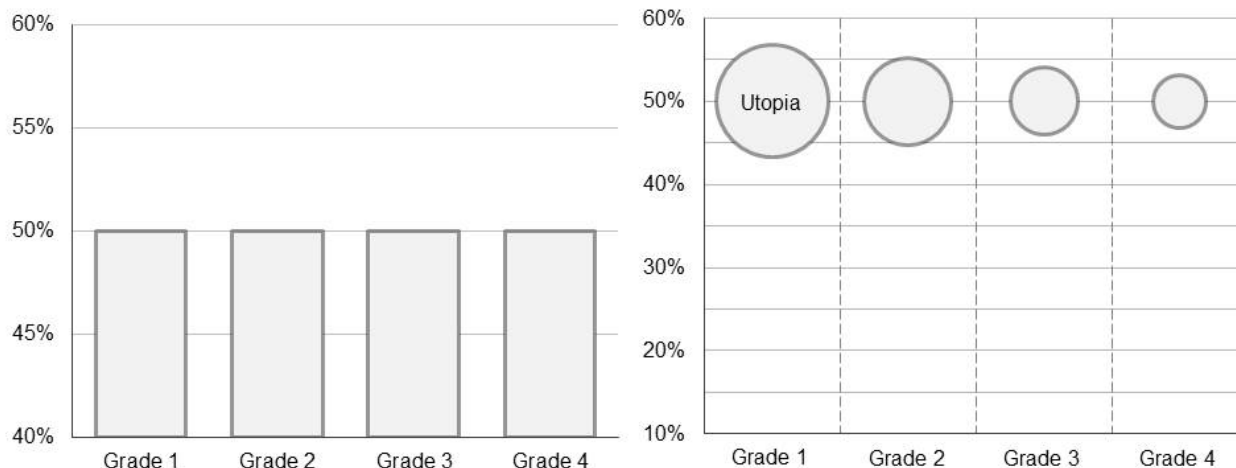


Figure 2: The University of Utopia, a) left, pipeline plot of percentage female staff at each grade, b) right, corresponding 'bubble' plot of the same data but where the size of bubble indicates the *number* of female staff at each grade.

Table 1: Data for construction of hypothetical *University of Utopia*. n = number of people at junior grade in department. p = progression rate, percentage of staff at one grade higher than previous grade, i = inequality index, percentage difference in the progression rate of men and women between each grade.

	Old department				New department			
	Size, n	% fem	Prgrss, p	Ineqilty, i	Size, n	% fem	Prgrss, p	Ineqilty, i
Utopia	50	50	0.6	0%	-	-	-	-
Scenario 1	50	50	0.6	0%	125	30%	1	0%
Scenario 2	50	50	0.6	0%	125	30%	0.4	10%
Scenario 3	50	50	0.6	5%	15	80%	1.2	10%

2.1.1 Scenario 1. Apparent leaks in the pipeline despite equal opportunities

The University of Utopia's management team decides that it would like to expand the university by opening a new department. The team chooses the new department to be a large scientific one ($n = 125$), in which—due to the inherent gender biases in the subject—it proves to be difficult to recruit equal numbers of men and women. Accordingly, the department is male dominated with only 30% of female staff in the junior grade. Also, because of the nature of the field, the promotion rates differ from that of the old department, where men and women still progress through the system with equal opportunity to reach the highest grade, but they do this at a slightly higher rate than the old department ($p = 1$). Notwithstanding these differences between old and new departments, management is determined to ensure that their gender equality practices are implemented in the new department such that the career prospects of both men and women are equal. Indeed, the new department succeeds in this aim

and has equal progression rates for men and women ($i = 0\%$). Therefore, across both old and new departments the university is a wholly equal opportunities university. After a few years the new department becomes established and the numbers of staff at each grade reach a steady state. At this stage the management team commissions an analysis of the fraction of women at each grade in the university. The data are presented to the team in the form of a leaky pipeline plot (Figure 3a). The team is dismayed and surprised when this plot shows a relative decline of women towards the higher grades. As a result, the management team devote new resource towards addressing the apparent problem and brings in new management policies to correct the data.

The truth of the matter in scenario 1, however, is that this is a hypothetical example of Simpson's reversal and that the management team was mistaken in thinking that there was some form of institutional inequality. There is no gender inequality in progression in either the old or the new department (shown by the 0% inequality index in Table 1 used to construct the data for the leaky pipeline

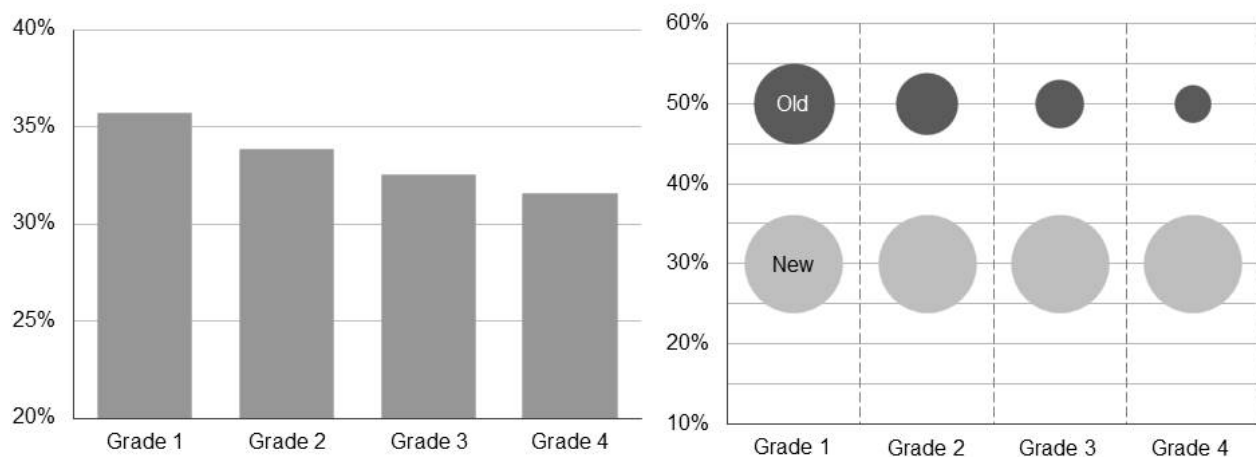


Figure 3: Scenario 1, a) left, leaky pipeline plot of the overall percentage of female staff at each grade, b) right, bubble plot of the same data shown for each department.

plots). The illusion created by the plot in Figure 3a is due to inter-departmental variation rather than intra-departmental variations, both of which are depicted in the bubble plot in Figure 2b where the number of female staff on each grade is shown by the size of the bubble. From this plot it is relatively easy to see that contributing 'weight' of each department to each grade changes moving from left to right, where the old department dominates the statistics at the junior grade (grade 1) and the new department dominates at the highest grade (grade 4). The resulting *overall* trend in Figure 3a therefore reports the variation *between* the departments (i.e. from one trend to the other) as one moves from left to right. This leads to the apparent illusion that there is a relative barrier to female progression in the university and, therefore, some gender inequality. Put another way around, despite their intuitive appeal, as this example shows, leaky pipeline plot can lead an evaluator to believe there is gender bias in progression when in fact there isn't.

2.1.2 Scenario 2. Gender inequality that is hidden by a leaky pipeline plot

In a different scenario, the University of Utopia's management team decide instead to create a new and large department of mathematics. Owing to the nature of the subject, there is a disparity in the levels of men and women recruited and the percent of female staff at the junior grade is only 30%. Also, the progression rates ($p = 0.4$) are lower than those in the old department. Unlike scenario 1, in this case the management team is lacking

in its efforts to ensure that the same gender equality practices in the old department are continued in the new one. Consequently a difference appears in the ability of men and women to progress from one grade to another ($i = 10\%$). After some time, following cases of reported gender inequality in the new department, the management team commissions an analysis in the form of an institutional leaky pipeline plot to assess the situation. Surprisingly, the percentage of women is shown to remain almost constant (Figure 4a) and claims of gender inequality are dismissed by the team as without basis.

The truth of the matter in scenario 2, however, is that there is a problem with gender inequality in the new department. This trend is clearly demonstrated in the bubble plot (Figure 4b) in which there is a downward sloping trend in the percentage of women at each grade for the new department. However, this trend is masked by the trend, which appears *between* the departments. The weight of the new department is highest at the lowest grade and the weight of the old department is highest at the top grade. The overall result is that the inter-departmental trend cancels out the intra-departmental trends with the net result that the overall statistics (Figure 4a) shows no variation in the percentage of female staff at each grade.

2.1.3 Scenario 3. Institution-wide gender inequality against women appearing as female advantage

Matters at the University of Utopia have taken a turn for the worse. The original 'completely fair' department

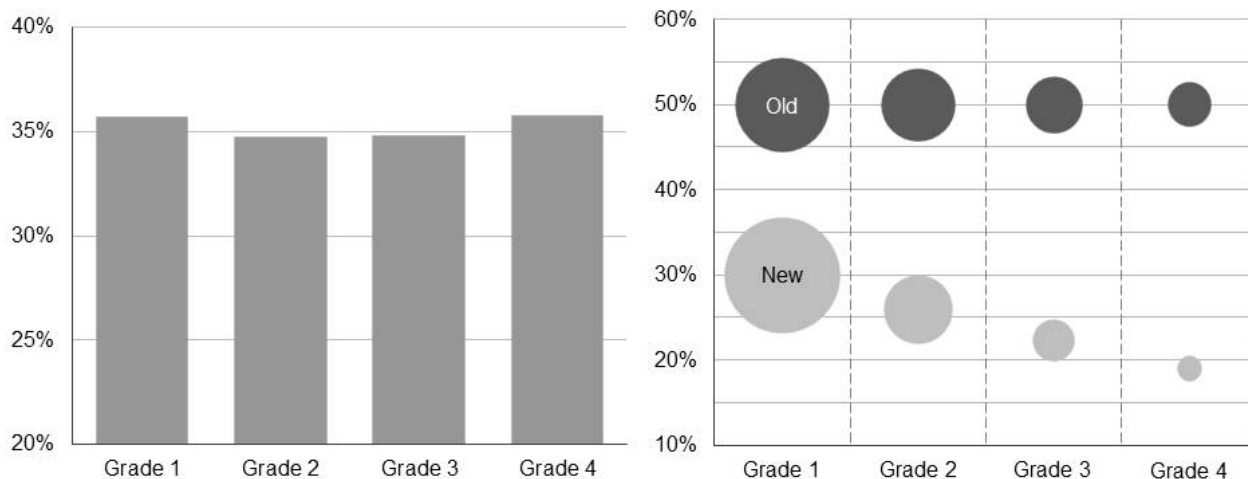


Figure 4: Scenario 2, a) left, leaky pipeline plot of the overall percentage of female staff at each grade, b) right, bubble plot of the same data shown for each department.

now has a degree of gender inequality against women ($i = 5\%$). A further small department has also opened which despite being female dominated ($\% \text{ male} = 20\%$) has significant levels of gender inequality in progression ($i = 10\%$). This new department differs from the original department by having higher levels of general progression through the grades such that there are more senior staff in this department than junior staff. The management team senses from qualitative surveys that there are levels of gender inequality disadvantaging women in both departments, and therefore commissions its usual leaky pipeline survey. The team is surprised when the data appear to show that women actually enjoy an advantage over men when it comes to promotion between grades (Figure 4a). It therefore concludes that the University needs to support more its male staff and brings in institutional policies to that effect.

Scenario 3 is the most extreme example of Simpson's paradox, and a salutary demonstration of how leaky pipeline plots can mislead. In it the overall leaky pipeline data actually show the opposite trend to the real situations in departments. Figure 5b reveals why this is the case. Despite the fact that both departments show declines in the percentage of women as they progress through the grades, the overall trend *between* the two departments is such that one department dominates the statistics at the lower grade and the other dominates at the higher grades. Therefore, progressing from left to right the trend essentially 'swaps' from one departmental trend to the other with the overall trend (Figure 5a) appearing to head in the opposite direction to the real situation. The result is that the overall leaky pipeline plot delivers the opposite trend to the one that exists—an extraordinary and counterintuitive result.

Scenarios 1,2 and 3 demonstrate how leaky pipeline data are vulnerable to misinterpretation through Simpson's paradox. Whether the scenarios come close to representing real situations is a matter of opinion, but since real situations will inevitably be more complex it is difficult to see how the paradox will be mitigated as more and more variables contribute to the overall data. This is the conclusion of other commentators on the subject who similarly conclude that it is not possible to use leaky pipeline data alone to establish inequality. Faced with this situation, managers and policymakers need to be aware of the problems associated with using leaky pipeline data, and need to be familiar with methods, which can help an evaluator spot whether the paradox is an important factor in their data. Such methods include the following:

1. Use professional statisticians to evaluate data,
2. Use disaggregated data where possible, in particular use plots like bubble plots or cluster analyses which can represent a third variable (i.e. number at each grade), although be aware that all aggregation levels are not free from the effects of conflating variables,
3. Use qualitative data and surveys to inform quantitative data,
4. Employ further statistical measures of significance, (Norton and Divine, 2015).
5. Also use other measures of inequality, like pay gap data.
6. Whenever percentages are used, also present absolute number data.

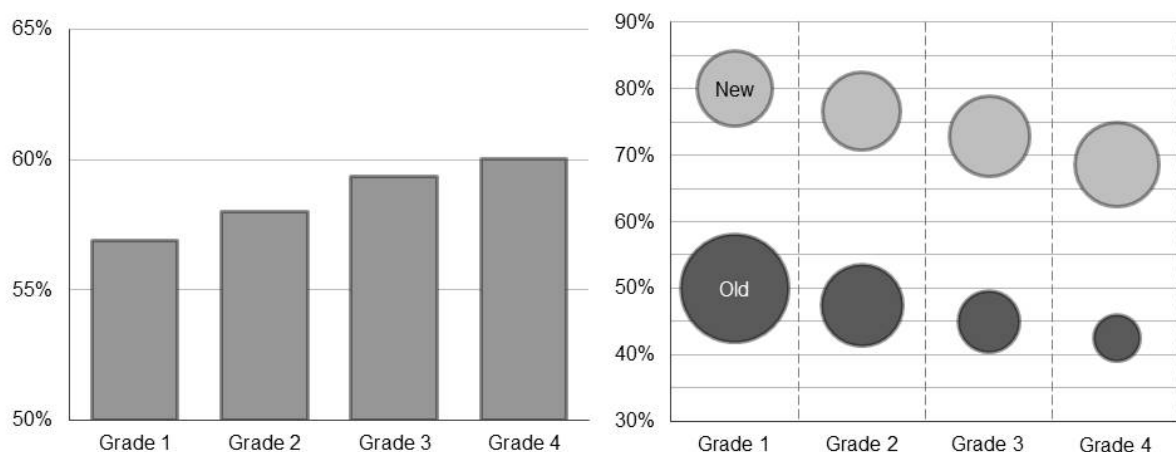


Figure 5: Scenario 3, a) left, leaky pipeline plot of the overall percentage of female staff at each grade, b) right, bubble plot of the same data shown for each department.

3 Conclusions

Leaky pipeline plots are widely used to measure gender (in)equality over a wide range of situations. The compelling nature of these plots means that they can become influential in policymaking, underpinning policy interventions and management decisions. The plots, however, like all quantitative estimators of behaviour are vulnerable to confounding variables, including inter-trend variations related to the relative size and progression rates of the contributing components. For example the increasing percentage of female staff at the higher grades shown in Figure 1 could be interpreted at first sight as less inequality compared to previous years, but it may also be due to the fact that the fraction of academic staff which are from clinical medicine departments increased significantly in the period 2010 to 2014 (19.3% of SET academics in 2010, and 21.3% of SET academics in 2014)—it is not possible to tell without more disaggregated data. One important manifestation of confounding variables is Simpson's paradox in which trends in overall progression data can be the *opposite* to the trends in the contributing components. While the existence of Simpson's paradox has been known for over a century and its hazards highlighted in gender equality practice, it remains widely misunderstood. Recent high profile examples have illustrated how even the most authoritative and in-depth studies can fall foul of its effects. It is evident, therefore, that awareness of the paradox needs to be raised. As part of this awareness raising, gender equality practitioners need to be equipped with the mental tools to overcome the highly counterintuitive nature of the paradox. We have shown herein that simple models can illustrate and exemplify the paradox within an understandable gender equality context. These models can then augment the recommendations and suggestions from several other commentators on how to avoid Simpson's paradox in not only interpreting gender equality data, but in all situations where fractional statistics are used.

References

- [1] Albers, C. J. (2015). Dutch research funding, gender bias, and Simpson's paradox. *Proceedings of the National Academy of Sciences*. 112 (50), E6828-E6829
- [2] Anderson, S. R., Auquier, A., Hauck, W. W., Oakes, D., Vandaele, W. & Weisberg, H. I. (2009). *Statistical methods for comparative studies: techniques for bias reduction*, John Wiley & Sons
- [3] Bickel, P. J., Hammel, E. A. & O'Connell, J. W. (1975) Sex bias in graduate admissions: Data from Berkeley. *Science*, 187, 398-404
- [4] Conroy, L., Watt, E. & Smith, W. Is there a 'Leaky Pipeline' for Women in Clinical Medical Physics in Canada? *World Congress on Medical Physics and Biomedical Engineering*, June 7-12, 2015, Toronto, Canada, 2015. Springer International Publishing., 1669-1672
- [5] Fenton, N., Neil, M. & Constantinou, A. (2015). Simpson's Paradox and the implications for medical trials. <http://www.eecs.qmul.ac.uk/~norman/papers/simpson.pdf> (accessed October 1st, 2016)
- [6] Jensen, J. L., Neeley, S., Hatch, J.B. & Piorczynski, T. 2015. Learning Scientific Reasoning Skills May be Key to Retention in Science Technology, Engineering, and Mathematics. *Journal of College Student Retention: Research, Theory and Practice*, 1521025115611616
- [7] McAllister, D., Juillerat, J. & Hunter, J. (2016). Funding: What stops women getting more grants? *Nature*, 529, 466-466
- [8] Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J. & Handelsmann, J. (2012). Science faculty's subtle gender biases favour male students. *Proceedings of the National Academy of Sciences*, 109(41), 16474-16479
- [9] Norton, H. J. & Divine, G. (2015). Simpson's paradox...and how to avoid it. *Significance*, 12(4), 40-43
- [10] Pearl, J. (2014). Comment: Understanding Simpson's Paradox. *The American Statistician*, 68(1), 8-13
- [11] Pearl, J. (2009). *Causality*. Cambridge University Press.
- [12] Shaik, F. & Fusulier, B. (2015). Academic Careers and Gender Inequality: Leaky Pipeline and Interrelated Phenomena in Seven European Countries. http://garciaproject.eu/wp-content/uploads/2015/12/GARCIA_working_paper_5.pdf (accessed October 15th, 2016)
- [13] Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, 13(2), 238-241
- [14] Valian, V. (1999). *Why so slow? The advancement of women*. MIT press
- [15] Van Der Lee, R. & Ellemers, N. (2015). Reply to Albers. Acceptance of empirical evidence for gender disparities in Dutch research funding. *Proceedings of the National Academy of Sciences*, 112(50), E6830
- [16] Verniers, C., Bonnot, V., Darnon, C., Dompnier, B., Martinot, D., Faniko, K., Lorenzi-Cioldi, F., Sarasin, O. & Mayor, E. (2015). How gender stereotypes of academic abilities contribute to the maintenance of gender hierarchy in higher education. In Faniko, K., Lorenzi-Cioldi, F., Sarasin, O., Mayor, E., (Eds.), *Gender and Social Hierarchies. Perspectives from Social Psychology*. London: Routledge
- [17] Volker, B. & Steenbeek, W. (2015). No evidence that gender contributes to personal research funding success in the Netherlands: A reaction to van der Lee and Ellemers. *Proceedings of the National Academy of Sciences*, 112(51), E7036-E7037
- [18] West, J. D., Jacquet, J., King, M. M., Corrells, S. J. & Bergstrom, C.T. (2013). The role of gender in scholarly authorship. *Plos ONE*, 8(7), e66212
- [19] Yule, G. U. 1903. Notes on the Theory of Association of Attributes in Statistics. *Biometrika*, 2(2), 121-134