CURRENT OPINION

Drug Safety 2007; 30 (10): 831-836
0114-5916/07/0010-0831/$44.95/0

A Case Study of a Graphical Misrepresentation

Drawing the Wrong Conclusions about the Measles, Mumps and Rubella Virus Vaccine

Anthony R. Cox[1,2] and *Harold Kirkham*[3]

1 West Midlands Centre for Adverse Drug Reactions, City Hospital, Sandwell and West Birmingham Hospitals NHS Trust, Birmingham, England
2 School of Pharmacy, School of Life and Health Sciences, Aston University, Birmingham, England
3 Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA

Abstract

Graphs have been used in attempts to show a relationship between the measles, mumps and rubella virus (MMR) vaccine and autism. We examine the topic of graphical representation of data in general, and one of these graphs in particular: the one that appeared in a 1999 letter to *The Lancet*. That graph combined data from England and from California, USA. The author alleged that this graph illustrated a rise in autism rates linked to the use of the MMR vaccine. By examining the presentation closely, we are able to show how this graph misrepresented the data used. We give advice for both authors and publishers in the use of such graphical treatments of data.

Graphs are often used in scientific and technical papers. They can make things easier to understand, clarify relationships, allow straightforward extrapolation and present a large amount of data concisely. However, they can also confuse and deceive; therefore, they should always be constructed carefully by the author and viewed carefully by the reader. The controversy over a link between the measles, mumps and rubella virus (MMR) vaccine and autism provides an example of why.

1. The 1999 *Lancet* Graph

The publication in 1998 of a paper by Wakefield et al.,[1] and a subsequent controversial press conference at which Wakefield called for suspension of the triple MMR vaccine,[2] led to a crisis in confidence about the MMR vaccine, which has had a detrimental effect on vaccination rates.[3] However, in 2004, the 1998 paper was retracted by ten of the authors,[4] and the editor of *The Lancet* stated that publication of the paper would have been handled differently if the full context in which the research had been done had been known.[5] Public confidence in the MMR vaccine has subsequently risen, with increased uptake of the MMR vaccine in 2005–6.[6]

The 1998 paper was not the only contribution to the MMR vaccine debate to appear in *The Lancet*. In September 1999, one of the original 1998 paper authors, Andrew Wakefield, published a letter in *The Lancet*,[7] containing a graph combining data from the Department of Development Services in California, USA,[8] with data from England obtained from a paper by Taylor et al.[9] (figure 1). The graph was used to allege that both sets of data illustrated a

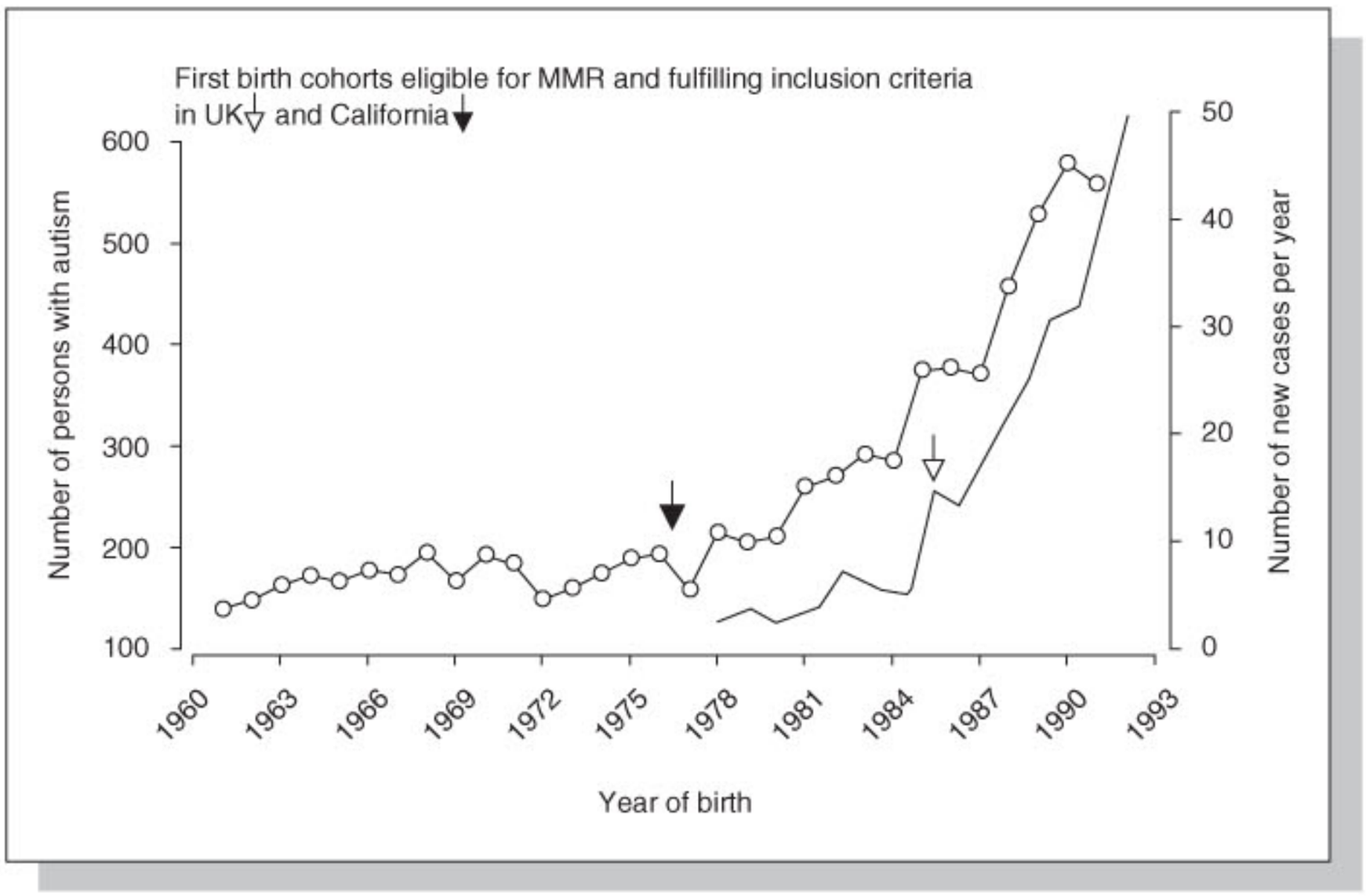

Fig. 1. The original graph published by *The Lancet* (reproduced from Wakefield.[7] Reprinted with permission from Elsevier [*The Lancet* 1999; 354: 949-50]).

rise in autism rates coinciding with the introduction of the MMR vaccine in each country.

The original caption accompanying the graph was as follows:

"***Temporal trends for autism in the USA (California*) and the UK (north-west London)***

In 1998 the expected numbers of newly diagnosed autistic children in California should have been 105–263 cases, according to DSM-IV; the actual figure was 1685 new cases. The temporal trend in north-west London is almost identical, although the rise is delayed by about 10 years. The two countries use the same diagnostic criteria. The sequential trends are consistent with the timing of introduction of MMR to both regions.

**Data from Department of Developmental Services, Sacramento, 1987–98 (www.dds.ca.gov).*"

It is not uncommon to use a time-series graph such as this to show correlation between two variables. Probably the example that springs most readily to mind is the similarity between planetary temperature variations and carbon dioxide levels shown by studies of ancient ice. However, it is often not good practice to use a time-series. Sometimes, the supposed association can be demonstrated only weakly. In any case, correlation does not imply causation.

In the case of the autism data of figure 1, the two data sets do not have a cause-and-effect relationship. They are presented by Wakefield[7] as if both curves represent an effect (autism) with a common cause (the MMR vaccine). Were that truly the case, the graphical approach would be both valid and useful. However, as we shall see, it is misleading to show the two curves in this way.

The California data were obtained from a 1999 graph produced by the Department of Developmental Services (DDS).[8] The authors of this diagram were at pains to point out in their report that their graph did *not* show how many people entered their system in a given year, but instead the number in the system *born in any given year*. The graph was a distribution of birth dates. They argued that the quality and type of information examined in their report were not suitable for measuring the incidence in the population of persons with autism. 'Incidence' in this context is a term of art meaning the number of people per unit of population diagnosed with a condition in a specified time, such as a month

or a year. Nor does the California report present prevalence data ('prevalence' is the number of people diagnosed with the condition per unit of population regardless of the date of diagnosis). That is why the vertical axis of the original California graph was labelled "Number of Enrolled Persons with Autism" as shown in figure 2. The authors of the California report were making the case to the legislature for increased departmental funding; theirs was not an academic study.

2. Root Cause

Instead of using a time-series graph, the authors of the California report should have used a bar chart. In fact, they should have used a specific kind of bar chart: a histogram. The use of a histogram, instead of a line graph, is required by the interaction between the bin width and the count. Were the bins to be made narrower (<1 year), the count in each bin would decrease. This is a characteristic of the histogram. The distribution of birth dates of people enrolled in a growing programme is not a time-series trend. Indeed, the data are neither continuous nor differentiable: the data do not represent a function of time. Given that the data show the count by years, the usual presentation is the population pyramid, a histogram with the bars horizontal.

Re-drawn as a population pyramid, the data from the original DDS graphic presents the Californian data in a different light. It is shown in figure 3 along with the English data, treated similarly. We discuss the English data later.

In figure 3, the meaning of the bars is unambiguous. The first horizontal bar for California can only be interpreted as the number of people 'in the system' who were 38 years old when the data were taken, and so on. The caption for the pyramid could have indicated that the data were analysed in 1998 for patient records up to 1992. It should also be pointed out that, in the period shown, the population of California grew by nearly a factor of 2. This fact alone must lead, other things being equal, to there being a higher number of young patients.

This population pyramid presentation is valid, and a line graph is not. However, this was not the graph published in California. Nor was it used by Taylor et al.[9] In both cases, a line graph was used instead.

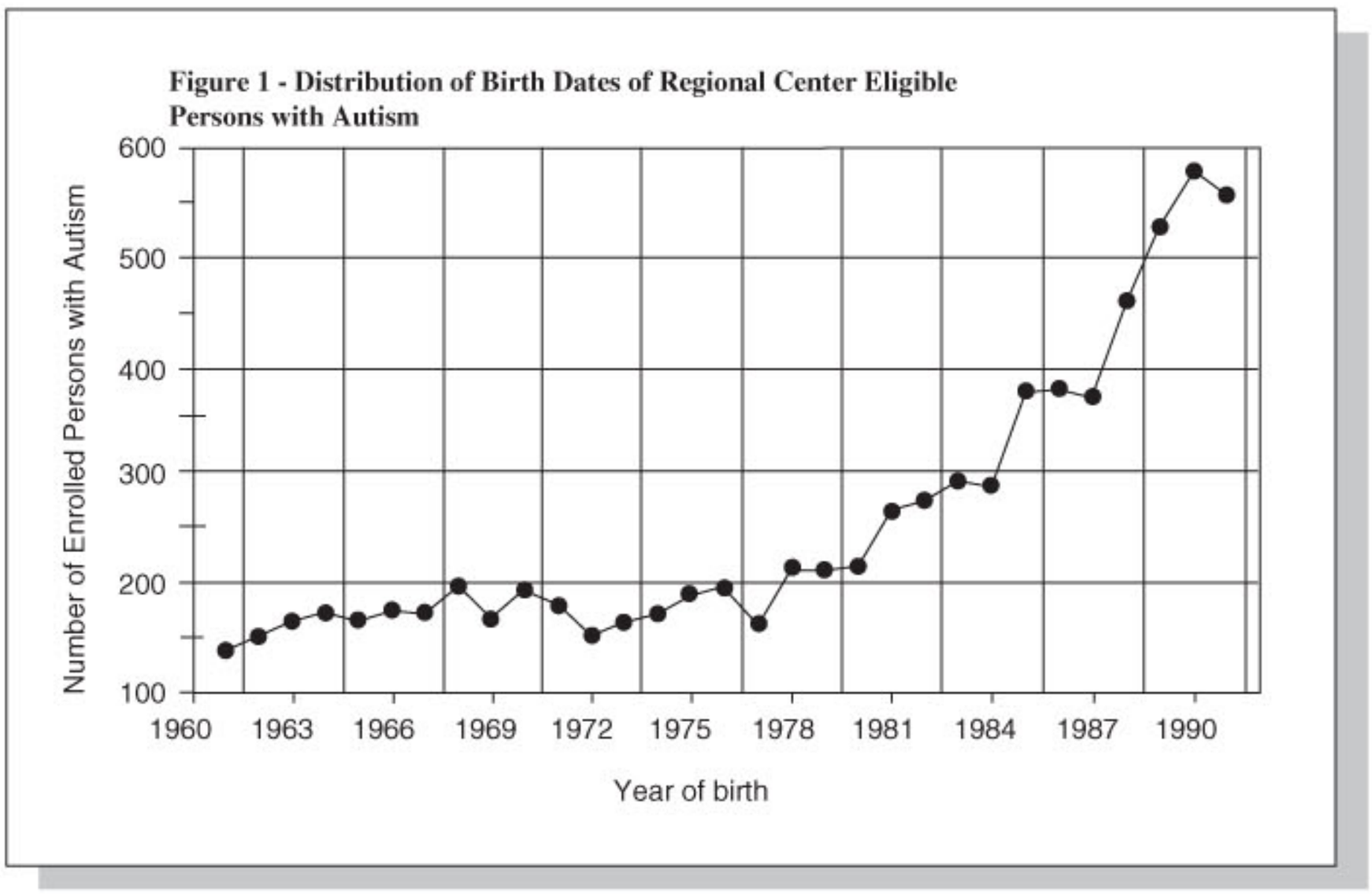

Fig. 2. The Californian chart (reproduced from Department of Developmental Services).[8]

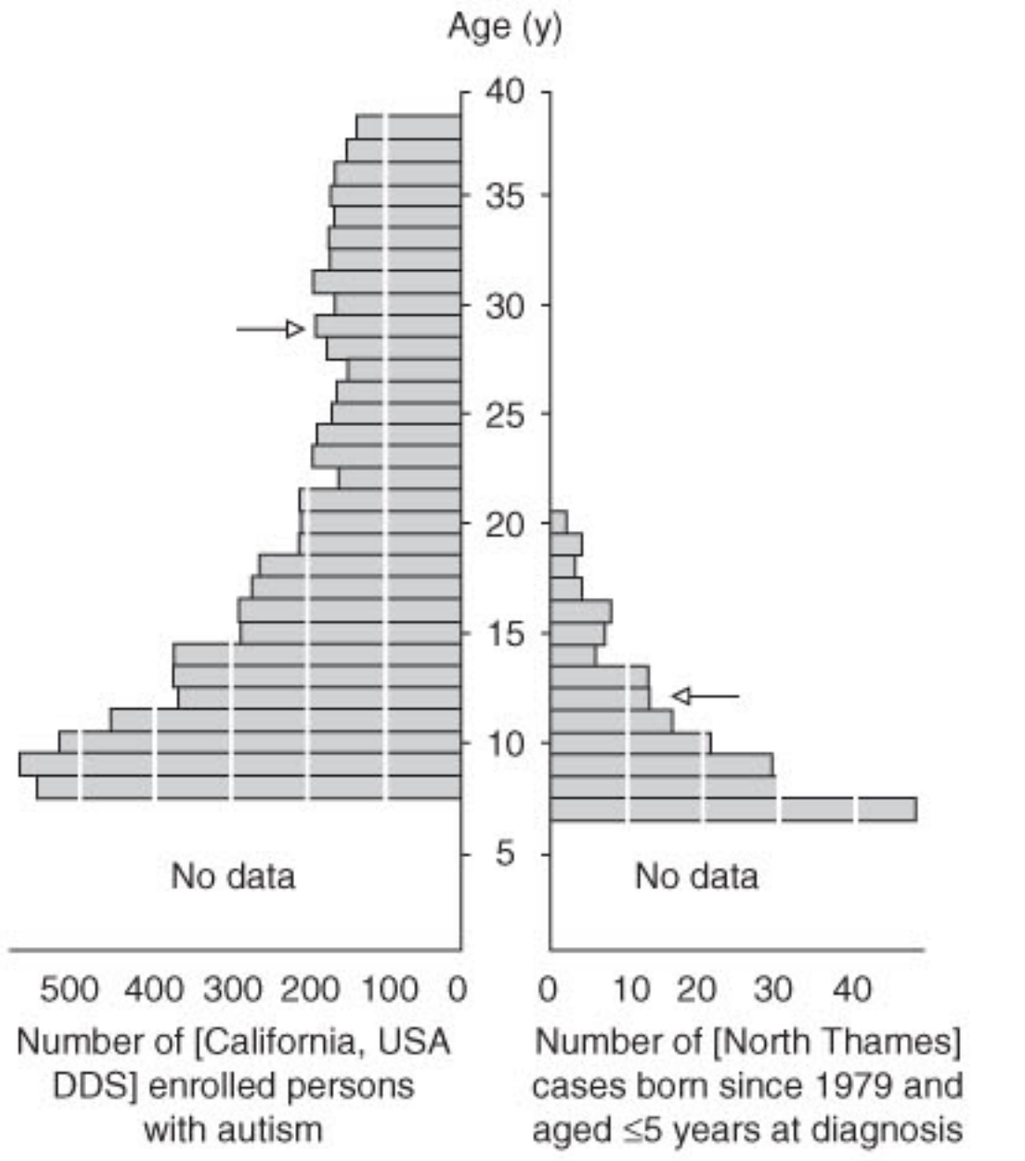

Fig. 3. The combined data presented as a population pyramid. Arrows indicate the year of the first birth cohort eligible for the measles, mumps and rubella virus vaccine. **DDS** = Department of Developmental Services.

3. Examination of the 1999 *Lancet* Graph

Taking the California graph[8] (figure 2) as the starting point, we can see that when the English data[9] were added to create the 1999 graph in *The Lancet*[7] (figure 1), a number of changes were introduced.

1. A new scale was added on the right, to be used with the added English data, labelled "Number of new cases per year".
2. The word "enrolled" was deleted from the label on the left, so it reads "Number of persons with autism".
3. The old figure caption, with its words about "Distribution of Birth Dates", is gone.
4. New words appear at the top of the graph, explaining that the arrows added to each curve indicate the "First birth cohorts [that were] eligible for MMR".
5. The data in each graph are a snapshot of the birth years of people in the system. We have no information about when they were diagnosed.

The California report[8] included all diagnosed cases of autism, without exclusion criteria. In contrast, Taylor et al.[9] obtained their data by selection from medical records. They selected patients who were in the records of eight North Thames health districts who were born since 1979 and before 1992, and who were aged ≤5 years at the time of diagnosis. The use of these criteria makes the data not directly comparable with the California data.

We note the following problems with the changes made to the graphs in order to merge the data sets:

- The new title for the graph, in boldface, is "Temporal trends for autism in the USA (California) and the UK (north-west London)". However, this is in defiance of the injunction given by authors of the original California graph not to use the information in this way.
- The new scale on the right starts at zero, whereas the suppressed zero of the original left scale has been retained. It is not valid in terms of the graphics to present one curve with a zero and the other with a zero suppressed. When the two things graphed are the same, there is at least an expectation by the reader that the offset is zero and the scale factor (at least when the data are normalised) is the same. It is not appropriate to increase the scale factor and change the offset of one of the graphs. In fact, the California numbers have a smaller dynamic range than the English results.
- The deletion of the word "enrolled" in the ordinate label is significant: it considerably broadens the meaning from the scope of the original, giving the impression that it was fair to compare the California data of enrolled children with the data from England, with its stricter inclusion criteria, supposedly representing "Number of new cases per year".
- The 1977 arrow purporting to show the start of the MMR vaccination programme in California is misleading in its precision. The combined MMR vaccine was licensed for use in the US in 1971;[10] therefore, the first eligible birth cohorts would have been those born a year or two before that. Throughout the 1970s, the combined MMR

vaccine replaced use of the individual measles, mumps and rubella vaccines.

- Locating the appropriate age for the arrows in the population pyramid makes a very different case. For several years after the introduction of the MMR vaccine in California, the number of people who (at some time) entered the DDS system remained more or less constant. This contrasts with the English data, where the numbers appear to be increasing even before MMR was introduced.

The number of changes associated with adding the English data is unusually large, and many of them are important in creating an impression in the mind of the reader. Of course, unless the reader has taken the trouble to examine the original graphs, none of the differences listed above will be obvious. As readers, we take it for granted that a citation is valid, that the author of the citation actually said what he is alleged to have said. It is a matter of trust between reader and author.

It is worth noting that a later publication by the California Department of Health Services shows no correlation between early childhood MMR immunisation rates in California and the numbers of children with autism enrolled in California's regional service centre system.[10]

4. Discussion

We believe there are lessons to be learned by both authors and journal editorial staff from this case study.

There are some general points for those wishing to use graphics:

- Choose carefully the kind of graph you use to show your data. Even if you select an apparently conventional kind of graph, be sure your selection is appropriate. Had the authors from the California DDS[8] and Taylor et al.[9] selected the histogram or the population pyramid, it would have been much harder to misinterpret their birth year distribution as a time series. The type of graph chosen by Wakefield[7] is understandable in view of the graphical choices made by the authors of the California report[8] and by Taylor et al.[9]
- Quote sources accurately. That applies both to verbal and graphical statements. In quoting verbal material, it is customary to show words left out (ellipsis) by printing dots or a long dash, and to show additions in square brackets. Changes to graphical material are not exempt from having such changes indicated.
- When referring to a graph, authors are under an ethical obligation to have read and understood the paper it was extracted from, and any surrounding explanatory text. In this case, the authors of the original California report clearly and repeatedly stated that the graph they had created did not show the incidence of autism in a given year or indicate any temporal trend. This word 'incidence' is chosen correctly; to show incidence, the numbers would have to show how many 'new cases' occurred in a given time per unit of population. The authors of the California report were not concerned with incidence, they were concerned with total numbers; theirs was a report to the legislature (for funding), not a science paper. One might note in passing that the study of Taylor et al.[9] does not show incidence, either.

There is also a lesson for medical journals. The alleged link between autism and the MMR vaccine has been refuted by both epidemiological studies[11,12] and virological studies;[13,14] the WHO Global Advisory Committee on Vaccines has also dismissed any link.[15]

However, the debate about the MMR vaccine continues outside the scientific community, particularly in tabloid newspapers in the UK. Although a lack of trust in the scientific consensus runs through these concerns, paradoxically the high reputation of journals is invoked in 'appeals to authority' by anti-vaccination campaigners. So, the respectability of *The Lancet* is invoked as a defence of the initial 1998 paper,[1] and the 1999 graph[7] can be invoked as further published evidence, in a peer-reviewed journal, for a link between the MMR vaccine and autism. It is therefore important that the publication

of such figures, even if they are correspondence items that are perhaps not normally subject to formal peer review, should be carried out with great care, especially in crucial areas of public health.

Statistical review policies of biomedical journals are not consistent, and it has been argued that improvements could be made in biomedical publishing;[16] we would add the recommendation that specific review of visual presentations of data should be also be made. Although it would have necessitated some investigation of the original sources, a review of the 1999 graph may well have influenced the decision about the suitability of the graph for publication, and prevented propagation of its erroneous message. Whether such specialist review is easily available is open to discussion, but given the controversial situation that existed when *The Lancet* letter was published, further scrutiny would have been justified.

Acknowledgements

We would like to thank Don Eckley, a retired pharmacist, for bringing us together to write this article, and Dr Patrick Waller, a consultant in pharmacoepidemiology, for his valuable advice relating to publication. We would also like to thank the anonymous reviewers of this article, whose thought-provoking comments improved the paper.

The authors came together specifically to write this article. Anthony R. Cox is the Pharmacovigilance Pharmacist at The West Midlands Centre for Adverse Drug Reactions in Birmingham and Teaching Fellow at The School of Pharmacy, Aston University, Birmingham, UK. Dr Harold Kirkham is a principal engineer at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA. He has an interest in the graphical treatment of data, and is in the process of writing a book on the topic.

Anthony R. Cox is also employed on a part-time basis at the Yellow Card Centre West Midlands, a regional education centre of the Medicines and Healthcare products Regulatory Agency (MHRA). The viewpoints expressed in this commentary are those of the authors and are not necessarily endorsed by the MHRA. Dr Harold Kirkham includes the 1999 *Lancet* graph discussed in this article as one of several case studies in the proposed book. No funding was received for the preparation of this review.

References

1. Wakefield AJ, Murch SH, Anthony A, et al. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. Lancet 1998; 351: 637-41
2. Fitzpatrick M. MMR and autism: what parents need to know. 1st ed. London: Routledge, 2004
3. Asaria P, MacMahon E. Measles in the United Kingdom: can we eradicate it by 2010? BMJ 2006; 333: 890-5
4. Murch SH, Anthony A, Casson DH, et al. Retraction of an interpretation. Lancet 2004; 363: 750
5. Horton R. The lessons of MMR. Lancet 2004; 363: 747-9
6. The Information Centre. Immunisation statistics England 2005–2006 [online]. Available from URL: http://www.ic.nhs.uk/pubs/immstats2005to2006 [Accessed 2007 Apr 4]
7. Wakefield AJ. MMR vaccination and autism. Lancet 1999; 354: 949-50
8. Department of Developmental Services. Changes in the populations of persons with autism and pervasive developmental disorders in California's Developmental Services System: 1987 through 1998 [online]. Available from URL: http://www.dds.ca.gov/autism/autism_main.cfm [Accessed 2007 Mar 26]
9. Taylor B, Miller E, Farringdon CP, et al. MMR vaccine and autism: no epidemiological evidence for a causal association. Lancet 1999; 353: 2026-9
10. Dales L, Hammer SJ, Smith NJ. Time trends in autism and in MMR immunization coverage in Calfornia. JAMA 2001; 285: 1183-5
11. Madsen KM, Hviid A, Vestergaard M, et al. A population-based study of measles, mumps, and rubella vaccination and autism. N Engl J Med 2002; 347 (19): 1477-82
12. Honda H, Shimizu Y, Rutter M. No effect of MMR withdrawal on the incidence of autism: a total population study. J Child Psychol Psychiatry 2005; 46 (6): 572-9
13. Afzal MA, Ozoemena LC, O'Hare A, et al. Absence of detectable measles virus genome sequence in blood of autistic children who have had their MMR vaccination during the routine childhood immunization schedule of UK. J Med Virol 2006; 78 (5): 623-30
14. D'Souza Y, Fombonne E, Ward BJ. No evidence of persisting measles virus in peripheral blood mononuclear cells from children with autistic spectrum disorder. Pediatrics 2006; 118 (4): 1164-675
15. Global Advisory Committee on Vaccines Safety. MMR and autism [online]. Available from URL: http://www.who.int/vaccine_safety/topics/mmr/mmr_autism/en/ [Accessed 2007 Apr 4]
16. Goodman SN, Altman DG, George SL. Statistical reviewing policies of medical journals, caveat lector? J Gen Intern Med 1998; 13: 753-6

Correspondence: *Anthony R. Cox*, West Midlands Centre for Adverse Drug Reactions, City Hospital, Sandwell and West Birmingham Hospitals NHS Trust, Dudley Road, Birmingham, B18 7QH, UK.
E-mail: a.r.cox@aston.ac.uk