

Do the print media “hype” genetic research? A comparison of newspaper stories and peer-reviewed research papers

Tania M. Bubela, Timothy A. Caulfield

β See related article page 1415

Abstract

Background: The public gets most of its information about genetic research from the media. It has been suggested that media representations may involve exaggeration, called “genohype.” To examine the accuracy and nature of media coverage of genetic research, we reviewed the reporting of single-gene discoveries and associated technologies in major daily newspapers in Canada, the United States, Great Britain and Australia.

Methods: We used neutral search terms to identify articles about gene discoveries and associated technologies hosted on the Dow Jones Interactive and Canadian NewsDisk databases from January 1995 to June 2001. We compared the contents, claims and conclusions of the scientific journal article with those of the associated newspaper article. Coders subjectively assigned the newspaper articles to 1 of 3 categories: moderately to highly exaggerated claims, slightly exaggerated claims or no exaggerated claims. We used classification tree software to identify the variables that contributed to the assignment of each newspaper article to 1 of the 3 categories: attention structure (positioning in the newspaper and length of the article), authorship, research topic, source of information other than the scientific paper, type and likelihood of risks and benefits, discussion of controversy, valuation tone (positive or negative), framing (e.g., description of research, celebration of progress, report of economic prospects or ethical perspective), technical accuracy (either omissions or errors that changed the description of the methods or interpretation of the results) and use of metaphors.

Results: We examined 627 newspaper articles reporting on 111 papers published in 24 scientific and medical journals. Only 11% of the newspaper articles were categorized as having moderately to highly exaggerated claims; the majority were categorized as having no claims (63%) or slightly exaggerated claims (26%). The classification analysis ranked the reporting of risks as the most important variable in determining the categorization of newspaper articles. Only 15% of the newspaper articles and 5% of the scientific journal articles discussed costs or risks, whereas 97% of the newspaper articles and 98% of the scientific journal articles discussed the likelihood of benefits of the research.

Interpretation: Our data suggest that the majority of newspaper articles accurately convey the results of and reflect the claims made in scientific journal articles. Our study also highlights an overemphasis on benefits and under-representation of risks in both scientific and newspaper articles. The cause and nature of this trend is uncertain.

CMAJ 2004;170(9):1399-407

The public gets most of its information about genetics from television, radio, magazines and newspapers.¹⁻⁴ It has been suggested that media representations of genetics are inaccurate or exaggerated and lead to a phenomenon that has been called “genohype,”⁵⁻⁷ the “hyped” portrayal of both the benefits and risks associated with genetic research and the application of genetic technologies. It has also been suggested that this sensationalization of genetics may have an adverse impact on the public’s ability to participate in policy discussions^{1,8} and on the utilization of genetic services because it creates inflated perceptions of the value of, for example, specific genetic tests.^{5,9}

Several studies and commentaries have suggested various degrees and sources of reporting inaccuracies. For example, institutional press releases may be incomplete (with risks and limits being underplayed),¹⁰ media stories about medications often include inadequate or incomplete information about risks and costs,¹¹ and abstracts from scientific meetings often receive a significant amount of media attention even though the validity and importance of the research has not “been established in the scientific community.”¹² Some have speculated about the general weaknesses of the popular media’s coverage of medical breakthroughs^{13,14} and their preference for positive over negative results.¹⁵

Surprisingly few systematic studies have examined the accuracy of media reporting in the context of genetics, most focusing on coverage of a single issue, such as sexual orientation^{16,17} or the discovery of susceptibility to breast and prostate cancer.^{18,19} The available data are instructive but not definitive.^{14,16-20}

We examined broadsheet newspaper coverage of genetic research in Canada, the United States, Great Britain and Australia, countries with a scientific community heavily involved in genetic research. We attempted to measure media “hype” and the factors that contribute to inaccuracy and exaggeration. In addition, we attempted to gain insight into the possible source (researchers, scientific journals or the media) and nature of any “hype.”

Methods

The search strategy identified gene discovery stories in the print media. We used neutral search terms — (“gene” or “genes” or “genetics”) and “discovery” and (“scientist” or “scientific”) and “research” — to identify suitable newspaper articles hosted on the Dow Jones Interactive and Canadian NewsDisk databases from January 1995 to June 2001. These databases offered the most complete collections of Canadian newspapers and included stories published before and after 2 landmark genetic events, the cloning of Dolly the sheep in 1997 and the completion of mapping of the human genome in 1999. Both of these events were highly publicized and probably increased the public’s interest in news coverage of genetics. We excluded articles on reproductive cloning because cloning was not the focus of this study and because the large number of newspaper articles on this subject had the potential to skew the results of the study. Although newspapers represent just one of many relevant media, we chose to study newspaper coverage because existing databases facilitated the comprehensive collection and analysis of relevant articles.

We then identified all newspaper stories that had as their source a paper in a peer-reviewed scientific journal. We used specific search terms to locate the scientific papers in MEDLINE or PubMed and then to identify all associated print media articles in the newspaper databases. The specific search terms included the name of the journal, the name of one of the researchers, the name of the institution where the research had been conducted, and the name of the gene, disease or genetic technology. We limited our study to all articles appearing in 26 newspapers from 4 countries. The broadsheet newspapers selected from the United States, Great Britain and Australia were chosen because they were national in distribution, had a large circulation, were of high quality, printed articles from one of the major newswire services and were roughly equivalent to the Canadian newspapers. We included most major Canadian local and national papers to ensure regional representation and thus captured most articles on gene discoveries and related technologies that were printed in that country. If an article appeared in more than one newspaper, it was coded only once; in such cases, priority was given to articles published in 1 of the 2 national newspapers, the *National Post* and the *Globe and Mail*.

We developed a coding frame, based on a previously published European media study,²¹ to compare the content of the scientific paper with the newspaper articles it generated. The coding frame consisted of a series of questions with standardized categorical responses on identifying information for the article; attention structure (positioning in the newspaper and length of the article); authorship; content; source of information other than the scientific paper; assessment of risk, benefits or controversy; valuation tone; framing; accuracy; and use of metaphors (see Appendix 1). All

questions except those on framing, authorship and attention structure were then repeated for the scientific paper.

The coders, all of whom had scientific backgrounds, were asked to subjectively assess both the technical accuracy of the newspaper article and whether the claims made in the newspaper article reflected those in the scientific journal article. For the latter, the coders assigned each newspaper article to 1 of 3 categories: moderately to highly exaggerated claims, slightly exaggerated claims and no exaggerated claims. We assumed that there were no technical errors in the scientific journal articles. A newspaper article was considered not to have been exaggerated if its claims had first been made in the corresponding scientific journal article.

We conducted a pilot study with 2 independent coders to test the coding frame. For the study itself, we used 3 independent coders who were not familiar with the project and who had a strong background in genetics and biotechnology. We held 2 moderated meetings early in the study so that the coders could discuss their interpretation of the coding frame.

To assess intercoder reliability, we asked each of the 3 coders to assess the same random selection of 84 (13% of the total) newspaper articles and their associated scientific papers. We calculated intraclass correlation coefficients (model 2)²² for questions in the coding frame with mutually exclusive response. In all cases the coefficient was greater than 0.75 (Table 1), which indicates good agreement.²²

We performed a classification tree analysis, using CART classification and regression tree software, version 4.0,^{23–25} to determine which variables from the coding frame contributed to the assignment of a newspaper article to 1 of the 3 categories of claims. This nonparametric statistical analysis has been used extensively for clinical risk assessment.^{25–28} The CART software uses specific criteria to split the data into 2 groups with greater homogeneity than the original group.^{24,25} Each group is then split into increasingly homogeneous groups until stopping criteria are reached and a set of terminal groups is produced. The variables are then ranked according to their importance as “splitting variables” (i.e., the relative contribution of each variable in assignment of the newspaper articles to the 3 categories).

Results

We examined 627 newspaper articles reporting on 111 papers from 24 scientific and medical journals. The most commonly cited scientific journals were *Science* (196 newspaper articles [31%]), *Nature* (120 [19%]), *Nature Genetics* (101 [16%]) and *Cell* (100 [16%]). Two hundred and seventy-six (44%) of the articles appeared in the A section of the newspaper, and 142 (23%) appeared on the first page.

Twenty percent (127) of the newspaper articles were written by clearly identifiable, in-house specialist science and technology or medical journalists, whereas 22% (138) were identified as being from a newswire service; the latter were almost exclusively brief announcements of a gene discovery with little or no analysis.

In 95% of the newspaper articles (591/625), the main source of information cited was the scientific paper or its authors (see Appendix 1). This is not surprising, given our

search strategy. Only 2 newspaper articles cited press releases as their source. Public sector research scientists (those perceived to be associated with a public sector employer, such as a university) were interviewed or cited in 90% of the newspaper articles (563/626), and private sector scientists in 5% (33/626). Opinions outside the research community or biotechnology sector were sought in only 8% (50/626) of newspaper articles, and these opinions were considered the main “voice” in only 2% (14/626) of the newspaper articles. Ninety percent of the newspaper articles (564/627) were framed as either a description of the basic research or as a celebration of progress or new research developments (Appendix 1). Most of the articles (483 [77%]) were not framed as a controversy; only 14% (85) were framed as a balanced controversy and 9% (59) as an imbalanced controversy.

The majority of newspaper articles were categorized as having no exaggerated claims (393 [63%]) or slightly exaggerated claims (165 [26%]). Only 11% (69) of the newspaper articles were categorized as having moderately to highly exaggerated claims (Table 1). Similarly, 82% (513) of the articles had no significant technical or scientific errors, and 18% (112) had 1 to 3 such errors. Only 2 of the

newspaper articles contained more than 3 significant errors. Canadian newspapers, both in general and in comparison with those in other countries, had low levels of exaggerated claims in their reporting (Fig. 1).

Forty-six (41%) of the 111 scientific papers reported research on human genetics, 40 (36%) used a nonhuman model, and 25 (23%) used both (Table 1). However, 87% (97/111) of the scientific papers and 98% (612/627) of the newspaper articles extrapolated the results to humans, regardless of whether the research model was human.

Beginning with the entire sample, CART identified the variable that best split the data into 2 subgroups with greater homogeneity than the whole and then repeated the process to create a classification tree (see Appendix 2). The relative contribution of the variables (their importance) as primary splitting variables was used in determining the category of exaggerated claims for the newspaper articles (Table 2). The misclassification cost for the learn data indicated a good fit of the model to the data: 0.25 for moderately to highly exaggerated claims, 0.24 for slightly exaggerated claims and 0.13 for no exaggerated claims.

The most important variable in determining the categorization of newspaper articles was the likelihood of risks or

Table 1: Questions on the coding frame comparing 627 newspaper articles with 111 scientific journal articles

Question and possible answers	No. (and %) of articles	Intraclass correlation coefficient*
<i>What research model does the scientific journal article use?</i>		
Human	46 (41)	0.98
Nonhuman	40 (36)	
Both	25 (23)	
<i>Does the scientific journal article discuss the research in terms of humans?</i>		
Yes	97 (87)	1.0
No	14 (13)	
<i>Does the newspaper article discuss a nonhuman model?</i>		
Yes	338 (54)	0.87
No	289 (46)	
<i>Does the newspaper article discuss the research in terms of humans?</i>		
Yes	612 (98)	1.0
No	15 (2)	
<i>Are there any significant technical or scientific errors in the newspaper article?</i>		
None	513 (82)	0.78
1–3	112 (18)	
> 3	2 (< 1)	
<i>Overall, do the main claims made in the newspaper article reflect the research findings?</i>		
Moderately to highly exaggerated claims	69 (11)	0.76
Slightly exaggerated claims	165 (26)	
No exaggerated claims	393 (63)	

*A measure of inter-coder reliability of the 3 coders.

costs being reported in the newspaper article, and the fourth most important variable was whether risks or costs were reported in the scientific paper (Table 2). Appendix 1 shows why mention of risk may be an important variable. Only 15% of the newspaper articles and just 5% of the scientific journal articles discussed costs or risks. Of the 94 newspaper articles that reported risks, 67 (71%) quantified the risk as very unlikely to somewhat likely, and 25 (27%) quantified the risk as very likely to already present. In the 6 scientific journal articles that mentioned risk, the assessment of risk ranged from very unlikely to very likely. In contrast, 97% of the newspaper articles and 98% of the scientific journal articles discussed the likelihood of benefits arising from the research (Appendix 1); the type and likelihood of benefit were therefore identified as less important in the CART analysis (Table 2). About 70% of the newspaper and scientific journal articles quantified these benefits as very likely or certain.

Also important were the themes of the scientific paper and the associated newspaper articles (Table 2). Newspa-

per articles on behavioural genetics (e.g., genes for sexual orientation, alcoholism, mental illness or criminality), genetically modified organisms (e.g., “glow-in-the-dark” rhesus monkeys or genetically modified killer pox virus), longevity (e.g., clock genes) or reproductive technologies (e.g., ooplasmic transplantation) were more likely to include moderately to highly exaggerated claims. Newspaper articles on diseases such as obesity were also more likely to include exaggerated claims, whereas articles on life-threatening and prevalent diseases such as cancer, stroke and heart disease did not include exaggerated claims (data not shown).

Our data may also indicate a more subtle form of media hype, in terms of what research newspapers choose to cover. Behavioural genetics and neurogenetics were the subject of 16% of the newspaper articles. A search of PubMed on May 30, 2003, with the term “genetics” yielded 1 175 855 hits, and searches with the terms “behavioural genetics” and “neurogenetics” yielded a total of 3587 hits (less than 1% of the hits for “genetics”). Al-

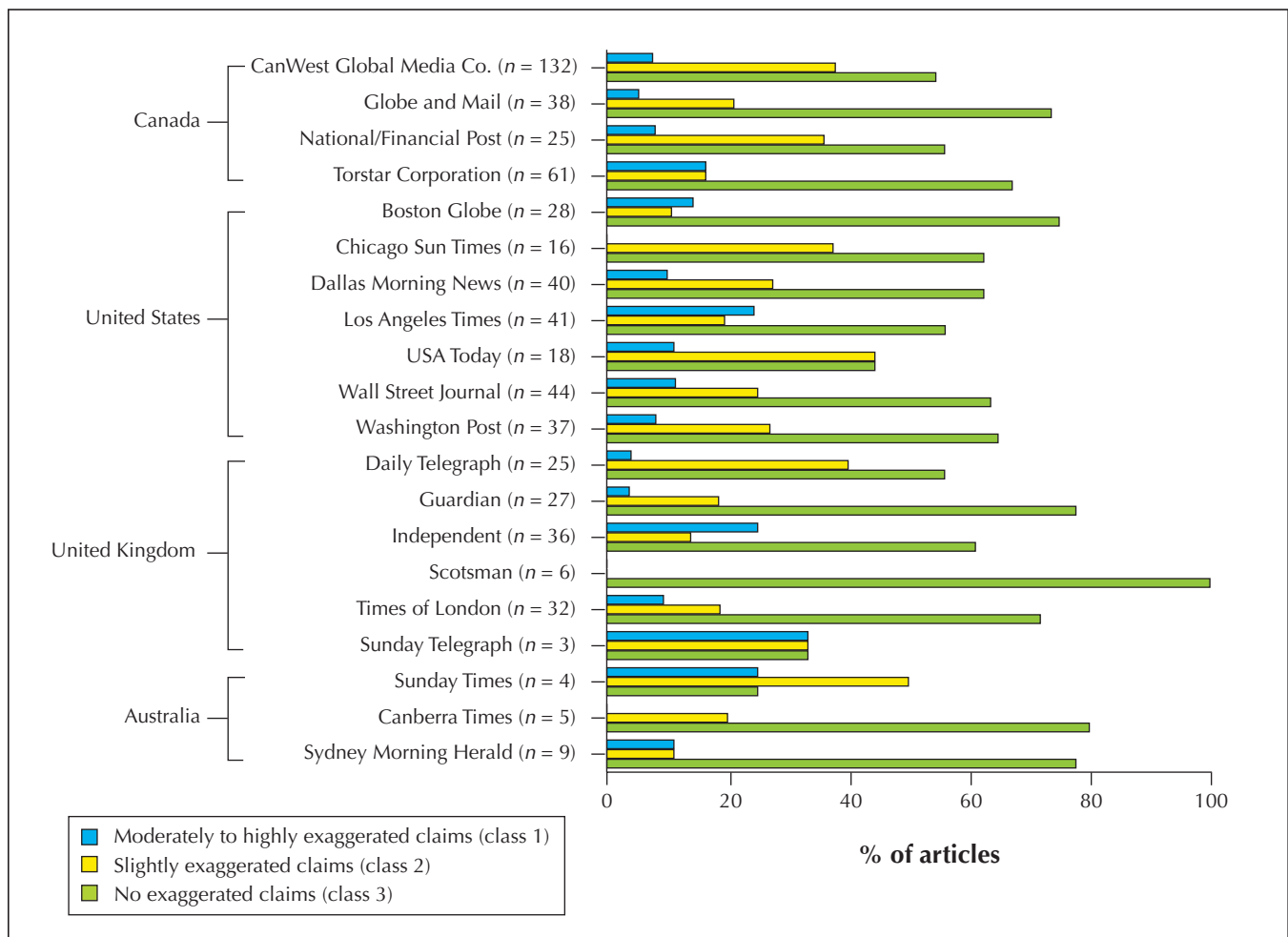


Fig. 1: Degree of exaggeration of claims in newspaper articles about genetics published in Canada, the United States, the United Kingdom and Australia, relative to their sources (scientific journal articles). The *n* values represent the number of articles from each newspaper.

though the latter is probably an underrepresentation of primary research papers in these fields and the former an overrepresentation of all research in genetics, the magnitude of the difference suggests that behavioural genetics is receiving a disproportionately large amount of coverage in the media.

Interpretation

Our data suggest that most newspaper articles accurately convey the results of and reflect the claims made in scientific journal articles. This high degree of conformity is consistent with other research results. For instance, Wilkes and Kravitz²⁹ found that most first authors (as listed in the byline) had a generally positive experience with the media, and 86% rated coverage of their scientific studies as accurate; only 3% called the coverage inaccurate. This level of accuracy corresponds well with our finding that most newspaper articles had no exaggerated claims (63%) or only slightly exaggerated claims (26%). However, the media do seem to over-emphasize particular topics, such as behavioural genetics. The high profile of these types of stories may be one reason for the perception that newspaper stories are often hyped. In other words, although we found that only 11% of the newspaper articles had moderately or highly exaggerated claims, these few stories might have a significant impact on public perceptions. As noted by Ransohoff and Ransohoff,⁸ “While most scientific writing is done well, the current level of exaggeration, even if infrequent, may discredit good reporting.”

In our study the 2 most important variables in determining the category of exaggeration were the reporting of risks and costs and the topic of the research. Our study also highlights an underreporting of risks and an over-emphasis on benefits. A small proportion of the newspaper articles (15%) and an even smaller proportion of the scientific journal articles (5%) dealt with risks. Again, this finding is consistent with existing data. For example, Moynihan and associates¹¹ found that only 15% of 207 news stories on drugs used for disease prevention presented both relative and absolute benefits. Similarly, in a study of Canadian newspaper coverage of new prescription

drugs, all articles mentioned some benefit, but the majority made no mention of side effects or harms.³⁰ Our study corroborates concern about the lack of reporting of risk, in both the scientific literature and newspaper coverage. In the rare instances that risk is reported in either type of article, there is a greater likelihood of exaggeration. In contrast, benefits are emphasized in both scientific and newspaper articles. This trend may contribute to a general hyping of genetic research, potentially inflating the expectations of the general public and special interest groups such as patient groups and investors.

Our study had a number of limitations. First, we surveyed only the print media and limited our analysis to the more respected newspaper publications, which do not necessarily have circulation levels as high as some tabloids. Second, our study was limited to newspaper stories

Table 2: Relative contribution of variables as primary splitting variables in a classification tree analysis to categorize newspaper articles on the basis of claims made*

Rank	Variable	Variable importance score, %†
1	Likelihood of risks or costs being reported in the newspaper article	100.0
2	Secondary theme reported in the newspaper article	87.8
3	Year of publication	64.5
4	Likelihood of risks or costs being reported in the scientific article	59.9
5	Secondary theme reported in the scientific article	50.9
6	Primary theme reported in the newspaper article	49.7
7	Primary theme reported in the scientific article	41.8
8	Main voice in the newspaper article	36.5
9	Type of risks or costs reported in the scientific article	24.0
10	Likelihood of main benefit being reported in the newspaper article	23.2
11	Type of main benefit reported in the newspaper article	22.1
12	Type of risks or costs reported in the newspaper article	13.5
13	Type of author of the newspaper article	11.2
14	Size of the newspaper article	10.7
15	Framing‡ of the newspaper article	10.2
16	Type of risks or costs reported in the scientific article	9.4
17	Person, organization or publication cited as the primary source of the information	7.4
18	Other sources cited or interviewed as authorities on the reported research	6.5
19	Type of main benefit reported in the scientific article	5.6
20	Presence of a negative valuation tone in the newspaper article	2.6
21	Presence of a negative valuation tone in the scientific article	2.3

*Categories: moderately to highly exaggerated claims, slightly exaggerated claims or no exaggerated claims.

†A higher variable importance score indicates a greater contribution to assignment of newspaper articles to 1 of the 3 categories, as determined by the classification and regression tree (CART) software.^{23–25} CART assigned a value of 100% to the highest-ranking variable and ranked all other variables relative to that variable. For example, according to the CART analysis, the secondary theme reported in the scientific article (variable importance score of 50.9%) and the primary theme reported in the newspaper article (49.7%), the fifth and sixth ranked variables, were equally important, but each had only about half the influence of the highest-ranking variable, likelihood of risks or costs being reported in the newspaper article (variable importance score of 100%).

‡Description or celebration of progress, economics or ethics.

that were directly related to peer-reviewed articles and published abstracts. As a result, stories flowing from other sources, such as abstracts from scientific meetings, which might never be published,¹² may be underrepresented. Also, all of our coders had strong scientific backgrounds, which undoubtedly brought a certain perspective to the interpretation of both the newspaper stories and the scientific articles.

Nevertheless, our data indicate that journalists may not always be the primary source of exaggerated claims. Although more research is required to confirm the nature and cause of this trend (e.g., investigating the editorial practices of the top journals and analyzing other media such as television and the Internet), a reasonable interpretation is that the media, scientific journals and the scientific community at large may be inadvertent “complicit collaborators” in the subtle hyping of science stories.⁸ Our data also raise interesting questions about how stories are selected and edited by the top scientific journals (e.g., the possibility of selection bias) and how the research community may “sell” science to the public, to scientific and lay publications, and to research funders.^{31–34}

This article has been peer reviewed.

From the Health Law Institute (Bubela, Caulfield), the Faculty of Law and the Faculty of Medicine and Dentistry (Caulfield), University of Alberta, Edmonton, Alta. Timothy Caulfield is Canada Research Chair in Health Law and Policy and is Research Director at the Health Law Institute.

Competing interests: None declared.

Contributors: Both authors contributed to the study conception and design and the data analysis and interpretation. Both drafted and revised the article and approved the final version.

Acknowledgements: We thank Edna Einsiedel and Usher Fleising for study design; Tim Karels for statistical advice; David Hik for statistical advice and editorial comments; Jordan Ward, Christine Caulfield and Qusair Mohammedbai for coding; and Nina Hawkins at the Health Law Institute for administrative assistance.

This study was supported by Genome Prairie, the Social Sciences and Humanities Research Council of Canada, the Alberta Heritage Foundation for Medical Research and the Stem Cell Network.

References

- Geller G, Bernhardt BA, Holtzman NA. The media and public reaction to genetic research. *JAMA* 2002;287:773.
- National Science Foundation, Division of Science Resources Statistics. *Science and engineering indicators 2002*. Arlington (VA): The Foundation; 2002 Apr. Available: www.nsf.gov/sbe/srs/seind02/pdfstart.htm (accessed 2003 Dec 3).
- Science and the public: a review of science communication and public attitudes to science in Britain*. London (UK): Wellcome Trust; Office of Science Technology; 2000. Available: www.wellcome.ac.uk/en/images/sciencepublic_3391.pdf (accessed 2004 Mar 26).
- Biotechnology Australia. *Biotechnology public awareness survey: final report*. Sydney: Millward Brown Australia; 2001. Available: www.biotechnology.gov.au/library/content_library/BA_MBsurveyresults.pdf (accessed 2004 Feb 20).
- Caulfield T. Underwhelmed: hyperbole, regulatory policy and the genetic revolution. *McGill Law J* 2000;45:437–60.
- Holtzman NA. Are genetic tests adequately regulated? [editorial]. *Science* 1999;286:409.
- Fleising U. In search of genohype: a content analysis of biotechnology company documents. *New Genet Soc* 2001;20:239–54.
- Ransohoff DF, Ransohoff RM. Sensationalism in the media: when scientists and journalists may be complicit collaborators. *Eff Clin Pract* 2001;4:185–8.
- Melzer D, Zimmern R. Genetics and medicalisation [editorial]. *BMJ* 2002;324:863–4.
- Woloshin S, Schwartz LM. Press releases: translating research into news. *JAMA* 2002;287:2856–8.
- Moynihan R, Bero CL, Ross-Degnan D, Henry D, Lee K, Watkins J, et al. Coverage by the news media of the benefits and risks of medications. *N Engl J Med* 2000;342:1645–50.
- Schwartz L, Woloshin S, Baczek L. Media coverage of scientific meetings. *JAMA* 2002;287:2859–63.
- Sibbison J. Covering medical breakthroughs. *Columbia J Rev* 1988:36–9.
- Loo LK, Byrne JM, Hardin SB, Castro D, Fisher FP. Reporting medical information: Does the lay press get it right? [abstract]. *J Gen Intern Med* 1998;13(Suppl 1):60.
- Koren G, Klein N. Bias against negative studies in newspaper reports of medical research. *JAMA* 1991;266:1824–6.
- Conrad P, Markens S. Constructing the “gay gene” in the news: optimism and skepticism in the US and British press. *Health* 2001;5:373–400.
- Petersen A. The portrayal of research into genetic-based differences of sex and sexual orientation: a study of “popular” science journals, 1980 to 1997. *J Commun Inq* 1999;23:163–82.
- Bernhardt BA, Geller G, Tambor E, Mountcastle-Shah E, Mull JG, Holtzman NA. Analysis of media reports on the discovery of breast and prostate cancer susceptibility [abstract 268]. *Am J Hum Genet* 2000;67(Suppl 2):62.
- Henderson L, Kitzinger J. The human drama of genetics: “hard” and “soft” media representations of inherited breast cancer. *Social Health Illness* 1999;21:560–78.
- Condit C, Ofulue N, Sheedy KM. Determinism and mass-media portrayals of genetics. *Am J Hum Genet* 1998;62:979–84.
- Gaskell G, Bauer MW, editors. *Biotechnology 1996–2000: the years of controversy*. London: Science Museum; 2001.
- Portney LG, Watkins MP. *Foundations of clinical research: applications to practice*. 2nd ed. Upper Saddle River (NJ): Prentice-Hall; 2000.
- Steinberg D, Colla P. *CART — classification and regression trees*. San Diego: Salford Systems; 1997.
- Breiman L, Friedman JH, Olshen RA, Stone CJ. *Classification and regression trees*. Belmont (CA): Wadsworth International; 1984.
- CART product overview [on-line]. San Diego: Salford Systems. Available: www.salford-systems.com/products-cart.html (accessed 2004 Mar 3).
- Begg CB. Statistical methods in medical diagnosis. *Crit Rev Med Inform* 1986;1:1–22.
- Steadman HJE, Silver J, Monahan PS, Appelbaum PC, Robbins EP, Mulvey T, et al. A classification tree approach to the development of actuarial violence risk assessment tools. *Law Hum Behav* 2000;24:83–100.
- Huland H. Radical prostatectomy: opinions and issues. *Eur Urol* 2001;39:3–9.
- Wilkes MS, Kravitz RL. Medical researchers and the media: attitudes toward public dissemination of research. *JAMA* 1992;268:999–1003.
- Cassels A, Hughes MA, Cole C, Mintzes B, Lexchin J, McCormack JP. Drugs in the news: an analysis of Canadian newspaper coverage of new prescription drugs. *CMAJ* 2003;168(9):1133–7.
- Nelkin D. *Selling science: how the press covers science and technology*. New York: Freeman; 1995.
- Lawrence PA. The politics of publication. *Nature* 2003;422:259–61.
- Media studies for scientists [editorial]. *Nature* 2002;416:461.
- Condit C. What is “public opinion” about genetics? *Nature Rev Genet* 2001;2:811–5.

Correspondence to: Timothy A. Caulfield, Health Law Institute, Law Centre, University of Alberta, 91 University Campus NW, Edmonton AB T6G 2H5; fax 780 492-9575; tcaulfld@law.ualberta.ca

Appendix 1: Results of coding for newspaper and scientific journal articles

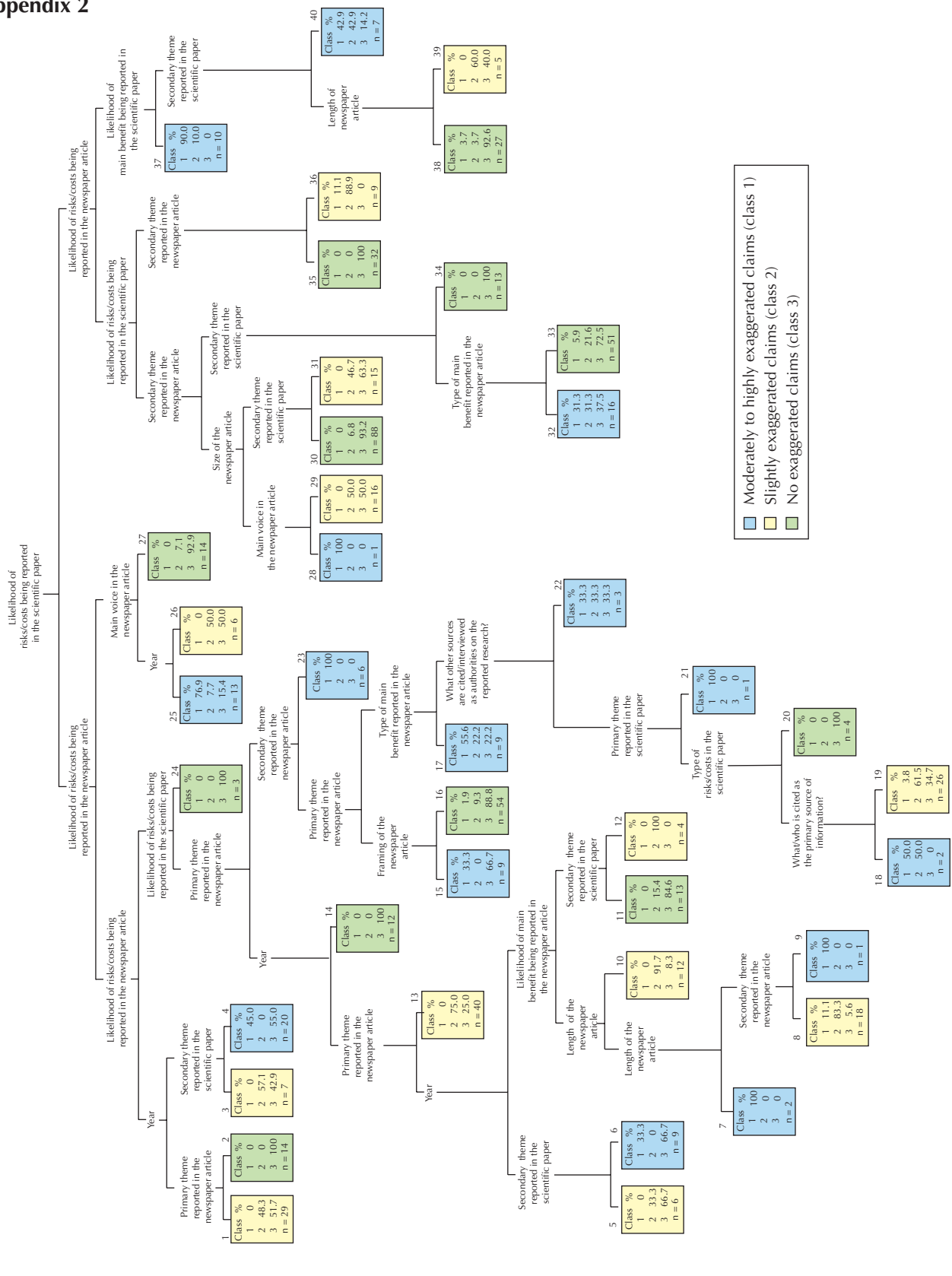
Category and option	No. (and %) of articles		Category and option	No. (and %) of articles	
	Newspaper*	Scientific journal		Newspaper*	Scientific journal
Main theme of article			Likelihood of main risk or costs		
<i>Science and medicine</i>			Very unlikely	1 (< 1)	1 (1)
Basic research	94 (15)	31 (28)	Somewhat unlikely	18 (3)	2 (2)
Disease or disease gene	298 (48)	56 (49)	Somewhat likely	48 (8)	1 (1)
Gene therapy	1 (< 1)	1 (1)	Very likely	10 (2)	2 (2)
Behavioural genetics or neurogenetics	98 (16)	14 (13)	Certain	7 (1)	
Genomics	8 (1)	2 (2)	Already present	8 (1)	
Pharmacogenetics	14 (2)		Not mentioned	532 (85)	105 (95)
Proteomics	18 (3)		Mentioned but not quantified	2 (< 1)	
Genetically modified animals	26 (4)	2 (2)	Total	626 (100)	111 (100)
Genetically modified humans	2 (< 1)		Negative valuation of biotechnology or genetic developments		
Cloning animals	12 (2)		Not to slightly critical, some discourse of concern	543 (87)	97 (87)
Stem cells	14 (2)	3 (3)	Moderately critical	71 (11)	13 (12)
Reproductive technology (not cloning)	6 (1)	1 (1)	Very to extremely critical, discourse of great concern	13 (2)	1 (1)
<i>Safety or risks</i>			Total	627 (100)	111 (100)
Health	2 (< 1)		Positive valuation of biotechnology or genetic developments		
<i>Other issues</i>			Not to slightly positive, discourse of promise	25 (4)	3 (3)
Diagnosis, genetic testing, predictive medicine (in adults or after birth)	4 (1)	1 (1)	Moderately positive	167 (27)	51 (46)
Economic prospects, opportunities	12 (2)		Very to extremely positive, discourse of great promise	435 (69)	57 (51)
Biopharmaceutical industry	8 (1)		Total	627 (100)	111 (100)
Legal or regulatory	6 (1)		Location of events		
Ethical	4 (1)		Canada	60 (10)	12 (11)
Total	627 (100)	111 (100)	United States	410 (65)	65 (59)
Is the article framed as a controversy?			United Kingdom	35 (6)	6 (5)
No	483 (77)	107 (96)	Australia	8 (1)	1 (1)
Yes, balanced	85 (14)	2 (2)	Other	114 (18)	27 (24)
Yes, imbalanced	59 (9)	2 (2)	Total	627 (100)	111 (100)
Total	627 (100)	111 (100)	Framing of story in newspaper article		NA
Type of main benefit			Descriptive or a celebration of progress	564 (90)	
Not mentioned	20 (3)	2 (2)	Economic prospect	10 (2)	
None (i.e., stated that there is no benefit)	1 (< 1)		Ethical	20 (3)	
Basic research	436 (70)	97 (87)	Risks after event	8 (1)	
Health	136 (22)	12 (11)	Nature versus nurture	4 (1)	
Economic	18 (3)		Profile or human interest story	21 (3)	
Legal	6 (1)		Total	627 (100)	
Social	3 (< 1)		Main voice (spokesperson, group or institution quoted or described in newspaper article)		NA
Political	2 (< 1)		Not applicable or unknown	5 (1)	
Moral, ethical	3 (< 1)		<i>Public sector</i>		
Total	625 (100)	111 (100)	Government research institution, scientist	91 (15)	
Likelihood of main benefit			University or hospital research scientists	472 (75)	
Very unlikely	1 (< 1)	1 (1)	The public, public opinion (e.g., survey)	2 (< 1)	
Somewhat unlikely	8 (1)	3 (3)	Media, published opinion	2 (< 1)	
Somewhat likely	96 (15)	17 (15)	Celebrity (sports, film, television)	3 (< 1)	
Very likely	256 (41)	41 (37)	<i>Private sector: business</i>		
Certain	198 (32)	38 (34)	Scientists in private laboratories	33 (5)	
Already present	42 (7)	9 (8)	Biotechnology company or spokesperson	7 (1)	
Not mentioned	20 (3)	2 (2)	CEO or upper management	4 (1)	
Mentioned but not quantified	2 (< 1)		Stock exchange	3 (< 1)	
Total	623 (100)	111 (100)	<i>Private sector: nonbusiness</i>		
Type of main risk			Patient groups or lobbies	3 (< 1)	
Not mentioned	532 (85)	105 (95)	Professional organizations	1 (< 1)	
Basic research	7 (1)		Total	626 (100)	
Health	48 (8)	5 (5)	Who or what is cited as main source of information for newspaper article?		NA
Economic	2 (< 1)		Scientific journal article or its authors	591 (95)	
Legal	1 (< 1)		Other scientists	8 (1)	
Social	5 (1)				
Political	1 (< 1)				
Moral, ethical	25 (4)	1 (1)			
Environmental, ecological	4 (1)				
Other	1 (< 1)				
Total	626 (100)	111 (100)			

Note: NA = not applicable, CEO = chief executive officer.

*For some variables, the total number of newspapers is slightly less than 627 because of coder error.

Continued on page 1406

Appendix 2



■ Moderately to highly exaggerated claims (class 1)
■ Slightly exaggerated claims (class 2)
■ No exaggerated claims (class 3)

Appendix 2: Classification tree showing the classification of newspaper articles into 1 of 3 categories: class 1 = moderately to highly exaggerated claims (relative to scientific article); class 2 = slightly exaggerated claims, class 3 = no exaggerated claims. The tree shown is the smallest tree within 1 standard error of the "relative cost" (i.e., misclassification rate) of the smallest tree determined by 10-fold cross-validation. Twoing for ranked data was the splitting method.^{24,25} The following paragraphs describe how the classification software functions (for an overview of the software used in this analysis, see www.salford-systems.com/products-cart.html) and details the left-most partitions of the tree.

The construction of classification and regression trees (CARTs) has become a common method for building statistical models from simple data. The results can be represented as binary trees, permitting simple graphic presentation, even when many variables are under consideration.

At each branching point or node of the decision tree is a binary question or statement (with a Yes or No answer) about some feature of the data set. The terminal groups or "leaves" of the tree represent the best split of the data, according to the "learn" or "training" data set. A terminal group may be a single member of some class (as in this analysis), a probability density function or a predicted mean value for a continuous variable. The basic CART algorithm is given a set of samples and instructed to find some variable that splits the data so as to maximize the differences (and minimize the similarity or "impurity") of the 2 partitions. This splitting of the data into groups is applied recursively until some stopping criterion is reached, such as a minimum number of samples in an individual partition (or branch of the tree).

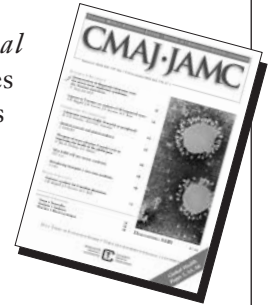
Because our purpose was to discriminate between 3 categories of newspaper articles (those with moderately to highly exaggerated claims, those with slightly exaggerated claims and those with no exaggerated claims), we used the classification tree analysis to determine the contribution of the descriptor variables listed in Table 2 to the assignment of the respective newspaper articles to the 3 categories.

The interpretation of this tree is straightforward. The left-hand partitions of the tree are described here in some detail, and the other branches can be interpreted in a similar fashion. The relative contribution of each variable to the overall shape of the tree is presented in Table 2 (as the variable importance score). At the top of the tree, all of the data are best divided into 2 groups according to the likelihood of risks or costs being reported in the scientific paper. The next node (branching point) to the left splits the data on the basis of the likelihood of risks or costs being reported in the newspaper article, and the same criterion applies to the third node. Year of publication is the variable that splits the remaining data into 2 penultimate groups. At the terminal split, the final groups (or "leaves" of the tree) are determined by the primary theme reported in the newspaper article. Node 1 (n = 29) corresponds to cases with slightly exaggerated claims in the newspaper article, whereas node 2 (n = 14) corresponds to cases with no exaggerated claims. Node 2 is "pure," in that all cases were assigned to class 3 (100%); however, node 1 is "mixed" because both class 2 (48.3%) and class 3 (51.7%) are represented. Further splitting would resolve these apparent impurities, but the CART pruning and stopping rules that we employed determined that this was the most appropriate terminal group.

CMAJ·JAMC

Manuscripts invited

The *Canadian Medical Association Journal* invites manuscript submissions from authors interested in wide exposure among Canadian physicians and around the world.



- Most widely circulated Canadian medical journal — 60 000 copies
- Highest readership of any scholarly journal in Canada
- Quality peer review
- Prompt editorial decision-making
- Fast-track review for articles of public health importance
- Indexed in MEDLINE, EMBASE/ Excerpta Medica, *Current Contents: Life Sciences*, *Current Contents: Clinical Medicine* and other abstracting/indexing services
- Full text available on eCMAJ: www.cmaj.ca
- Biweekly media coverage coordinated by CMA

Guidelines for authors are available on eCMAJ (www.cmaj.ca).

To discuss submission of a manuscript, contact:

John Hoey, MD
 Editor, CMAJ
john.hoey@cma.ca
 tel 800 663-7336 x2118 (in North America)
 tel 613 731-8610 x2118 (from other points)
 fax 613 565-2382

