Academic Writing Center

Academic Writing Workshop Booklet

Fall 2019

Katherine Willcox Özsarı

Academic Writing Center

Part I: Introduction to Academic Writing

A. General Approach to Academic Writing

Academic Writing as a Conversation

Starting with what "They Say"

> "Academic writing… calls upon writers not simply to express their own ideas, but to do so as a response to what others have said. …remember that you are entering a conversation and therefore need to start with 'what others are saying'… and then introduce your own ideas as a response. This is not to say that you must start with a detailed list of everyone who has written on your subject…"
>
> - *They Say, I Say* by Gerald Graff (pp. xvi, 20)

Exercise 1: In groups of 3-4, discuss the meaning of the above quote. What is Gerald Graff's recommendation? Do you think this is relevant to your field of study? How so? How should we approach research and writing? Do you have other general recommendations for academic writing in your or other departments? You can take notes below.

_____

_____

They Say, I Say

Remember that the purpose of all academic writing is the **exchange of ideas** among experts in order to answer important questions, test hypotheses, or solve problems. The common objective is to understand ourselves and the world. Scholars are not isolated. They often collaborate with others, and even single-author papers enter a **conversation** with other specialists all over the world.

Self-regulation in Academic Writing

Academic writing is self-regulating, making it a true community. Each new article is a response to research that preceded it and anticipates future research. You will be publishing most of your articles in **peer-reviewed journals**, in which submissions are reviewed by experts before publication. Researchers question, test, retest, and criticize each other's' works. Academic writers often anticipate criticism or challenges to their methods or findings. They explain possible weaknesses in a section at the end of the article that addresses limitations. This invites other researchers to strengthen results in future articles to make results more reliable.

How Can Academicians Improve Academic English?

Exercise 2: **What do you think?** In groups of 3-4, discuss what you think is the best way to improve academic English. Write your recommendation below and share with another group.

_____

_____

Exercise 3: **Do you agree?** Now read the following quotes by famous language acquisition scholars on how to improve a second language. Mark whether you agree or disagree with each.

Agree / Disagree

_____ / _____ Knowledge of physics is conscious knowledge; the physicist can expound and articulate it and convey it to others. In contrast, the other two systems [grammar i.e. mental representation, and common sense] are quite unconscious for the most part and beyond the bounds of introspective report. Grammar and common sense are acquired by virtually everyone, effortlessly, rapidly, in a uniform manner, merely by living in a community under minimal conditions of interaction, exposure and care. There need be no explicit teaching or training, and when the latter does take place, it has only marginal effects on the final state achieved. – Noam Chomsky

_____ / _____ The source of good writing style, the vocabulary, syntax and discourse structure of the written language, is reading. – Lee & Hsu

_____ / _____ The best way to improve in a foreign language is to do a great deal of comprehensible, interesting reading. – Mason

_____ / _____ The best way to improve your knowledge of a foreign language is to go and live among its speakers. The next best way is to read extensively in it. –Nuttal

_____ / _____ Even after puberty, the brain is elastic enough to internalize a second (or third) language basically in the same manner it picks up the first. – Ellidokuzoglu

_____ / _____ Many studies confirm that those who read more write better… it is reading, not instruction, that helps us develop a good writing style. – S. Krashen

_____ / _____ People acquiring a second language have the best chance for success through reading. -S. Krashen

_____ / _____ Free voluntary reading results in better reading comprehension, writing style, vocabulary, spelling and grammatical development. – S. Krashen

_____ / _____ There is no need for deliberate memorization; rather, firm knowledge of grammatical rules and a large vocabulary gradually emerge as language acquirers get more “comprehensible input,” aural or written language that is understood. -S. Krashen

_____ / _____ Picking up word meanings by reading is 10 times faster than intensive vocabulary instruction. -S. Krashen

_____ / _____ No meaningful support has [ever] been provided for the position that grammar should be taught. – M. Long

_____ / _____ Incidental learning of words during reading may be the easiest and single most powerful means of promoting large-scale vocabulary growth. – W. Nagy., R. Anderson & P. Herdman

Exercise 4: Based on the quotes above, discuss the following questions in groups of 3-4.:

1. How is learning science different than mastering a language?

According to the quotes: _____
_____

Does your group agree? Why/ Why not?

_____

_____

2. What is the fastest and most powerful way to improve grammar, vocabulary and writing?

According to the quotes: _____
_____

Does your group agree? Why/ Why not?

_____
_____

3. Based on the above quotes, would you like to revise your recommendation in **Exercise 3** (What is the best way to improve academic English)? If so, rewrite it below:

_____
_____

B. Introduction to Academic Writing

Becoming Familiar with Academic Discourse and Conventions

In order to become a good writer of academic prose, it is necessary to first become familiar with the conventions and vocabulary of academic discourse. The best way to do this is to read articles in the top journals in your field. **Becoming familiar with these conventions will help you connect with your reader once you begin writing.**

Conducting Research

Research often begins after you have come up with a research question or a statement of the problem to be investigated. However, your research question or thesis will likely not be clear until you have conducted preliminary research. Reading articles will help you understand what research has already been done (which will help you write your **literature review**), and what is missing in the research that needs to be studied (which will help you write the **gap in current research** and decide on your **thesis**). Reading extensively will also improve your writing and knowledge.

Resources for Research

Once you know your topic, problem, or general area of interest, you can begin "**reviewing the literature**" to get a better idea of what you should focus on. Reviewing the literature means looking at the research related to that topic.

In determining whether to use a source in your research, bear in mind **the four "R's**" of research sources: reputable, reliable, recent and relevant.

- **Reputable**: Reputable sources are usually associated with well-known organizations or acknowledged experts in their field. Peer-reviewed journals are a reputable source.
- **Reliable**: Information from reliable sources can be trusted as accurate and free of bias.
- **Recent**: Although currency is more important to some topics than others, recent information is generally superior to older information.
- **Relevant**: The information in relevant sources is directly related to your thesis and/or main points.

Databases

A **reliable source** for your research is an academic/scholarly **book** or **journal** (a small collection of articles), published by a well-known organization. These are **peer-reviewed** (approved by other academicians or experts in the field).

Fortunately, there are many **databases** in your field that have already done the job of choosing the best articles and books for you. Good academic **databases** are search engines full of academic **articles** from the best academic **journals**. For instance, **ScienceDirect** is a database that includes peer-reviewed and edited scientific and mathematical articles from top journals. **JStor**, on the other hand, is perfect for those studying the social sciences, like sociology or literature. **Proquest** has articles from most fields. Our library website lists databases by subject, and you can also ask your professor which one he or she uses. You may be familiar with **Google Scholar** (a database of academic articles and books) and **Google Books** (a database of books); those are great sources too, but they include a wider variety of subjects and source types, so they can be difficult to pick through.

You may also find hard copies of books and journals in the library. However, most of your research will likely be done online, since reputable journals publish articles on the internet. Note, though, that you have to be on campus to access many databases. If you want to read an article at home, you can first save it to your computer while on campus.

Recording Important Information as you Research

Keeping records during the research phase of writing allows you to read material efficiently as well as save time when you write your paper.

Record notes that include the following information:

- A direct quotation, summary or paraphrase of the main idea
- Complete names of author(s) and/or editor(s)
- Complete name of the book, journal, magazine, newspaper or website
- The title and full publication details, including date and edition, publisher
- The page numbers you consulted

Referencing software tools such as **Endnote**, Mendeley, and RefWorks can help you record and organize your research and, later on, cite and reference your research in your article. For tutorials and more information on how to use Endnote for free, contact the IzTech Library.

Planning Your Writing

Narrowing your Topic and Developing your Thesis

Your research and your advisor can help you narrow down your topic and main thesis or questions. You should get a feel for what topics are currently of interest in your field and what research questions have not yet been answered. Your topic should be narrow and specific and yet broad enough that you can tackle it in an article or dissertation.

Creating an outline/structure

Academic writing requires a clean, organized, well-structured outline. You must create a "scaffolding" that's attractive, inviting, and easy to navigate, without any unnecessary or confusing bits hanging out. A clear structure is not only more inviting, but also easier to navigate for your reader. An outline will also help you decide which information you really need from your sources, and where to place them in your paper. Before you even finish your research, you should take the time to create a detailed outline.

Conduct further research and experiments

Now that you have a narrower topic, a thesis or main idea, and a solid outline, conduct further research and experiments. Insert your notes beneath the appropriate section in your outline.

Integrating and Documenting Sources

Academic journals usually employ one documentation method consistently, using a set of established standards for citing and referencing sources. Academic journals in similar disciplines may not always use the same method. You should check the documentation method used by your intended journal before you write or send your article.

Integrating Sources

You typically integrate, or synthesize, your sources as part of the composing process. When you **summarize** a source or sources, you extract ideas that are directly relevant to your writing, expressing them in your own words. What distinguishes a summary from a **paraphrase** is that summaries are selective: they focus on main ideas. When you paraphrase, you include *all of the original, putting it in your own words.* Paraphrasing is reserved for very important detailed information. Whereas a summary condenses and is thus an efficient method for synthesizing material, a paraphrased passage is not usually shorter than the original – in fact, it may be longer. Paraphrase is used sparingly in scientific writing. Direct quotation, where you use exactly the same words as the original within quotation marks, is employed even more rarely in scientific writing (Henderson, 2012).

Documenting Sources

Documenting sources in a **list of references** is often the last stage of writing your paper, although you should keep track of your sources throughout the research process. Documenting sources serves several practical purposes: it enables a reader to distinguish between your ideas and someone else's, and it makes it possible for any reader to access the source itself to ensure its accuracy or focus on its content.

Plagiarism

Plagiarism is an extremely serious academic offence. You plagiarize if you use any material that is not your own – whether you quote directly, summarize, paraphrase, or refer to it in passing – without acknowledging it. You plagiarize if you use the exact words of the source and do not put them in quotation marks and if you follow the structure of the original too closely.

What kind of information must be cited?

You do not need to cite anything that falls under the category of *general knowledge.* If a typical reader is likely to know something, a citation may be unnecessary. If a fact or idea is *easily obtainable*, a citation may also be unnecessary. This will depend on the expertise level of your audience.

Voice and Style in Academic Writing

According to Henderson, academic prose is "**customized for an audience familiar with a given discipline's conventions and modes of discourse, its central ideas and its ways or presenting and analyzing them**." It is important to **have your intended audience in mind** when you are writing your dissertation or academic article. Not all readers will be specialists in your field. The readers of Academic journals vary from very knowledgeable readers to those with general knowledge. Biologists tend to read the journal *Cell*, while physicists are the main audience of *Communications in Mathematical Physics*. However, academicians from many fields subscribe to scientific journals like *Nature* or *Science.* Becoming familiar with the conventions of the type of journal in which you wish to publish can help you target your prose toward your audience **so that they can understand your research.**

Some people think of academic writing as dense, abstract writing so highly specialized that non-specialists cannot understand it. However, successful academic writing **should not confuse the reader**. Academic writers may use specialized diction or jargon. However, academic writing includes less ornamentation than literary writing. It is marked by **direct, straightforward prose with few modifiers** (adjectives and adverbs).

The voice in academic essays is generally **objective and analytical**, avoiding the expression of personal views. They use voice in specific ways to **assume distance from the study and avoid bias**. For example, writers may use passive constructions, in which the subject of the sentence is acted upon, rather than acting itself.

Passive vs. active voice: Student writers are often told to avoid the passive voice in their writing, because it often results in a weaker sentence. For example, instead of saying "The metal beams were corroded by saltwater" (passive), students may be encouraged to write "Saltwater corroded the metal beams" (active). The passive voice can be used to de-emphasize the subject (such as the researcher) or stress the object (such as that which is being studied). For instance, instead of writing "We studied the correlation between red meat consumption and cancer," scholars may prefer to say "The correlation between red meat consumption and cancer was studied."

However, it is becoming more common in science writing to employ the **active voice**, for the sake of simplicity. Sometimes it is much easier to understand a sentence written in the active voice. For example, some articles will use sentences similar to the following: "We then injected *smyd5*-MO into embryos and examined gross morphological phenotypes."

So, **how can you know whether to prefer active or passive voice in specific situations**? The best way is to read numerous articles in the top journals of your field to get a feel for academic language used in your research area.

Exercise 5: Below is an abstract from the journal *Child Development*. Abstracts precede many journal articles to give a brief summary of the article. **Underline the use of passive voice**, and **put a star next to the use of active voice.**

Using a genetic design of 234 six-year-old twins, this study examined (a) the contribution of genes and environment to social versus physical aggression, and (b) whether the correlation between social and physical aggression can be explained by similar genetic or environmental factors or by a directional link between the phenotypes. For social aggression, substantial (shared and unique) environmental effects but only weak genetic effects were found, for physical aggression, significant effects of genes and unique environment were found…

-M, Brendgen, G. Dionn e, A, Girard, M. Boivin, F Vitaro, & D. Perusse (2005), "Examining genetic and environmental effects on social aggression: a study of 6-yearold twins." Child Development, 76: 930-46

Follow-up Question: Does the abstract employ mostly active or passive voice? What is the effect or purpose of the voice used? Do you think abstracts usually employ active or passive voice? Why?

_____

_____

Different Types of Academic Writing

Research Proposals

Research proposals are usually written before your major research, but can also be written afterwards. The main purpose of the proposal is to convince a reader that the project you propose is **worth doing** and that you are the right person to do it.

Research proposals need (1) a description of what you are undertaking, (2) your methodology, and (3) the sources you have found useful and the sources you may use as your research continues. The proposal represents a *probable* plan, but your thesis and main points can be revised if necessary.

Dissertations / Theses

The organization and content of your thesis/dissertation will vary based on your discipline and topic of study. A thesis includes a deeper study of a research topic under expert supervision. A thesis discusses current as well as future research options and presents the research in a more detailed manner, with each component divided into one or more chapters. A thesis is usually written after a student completes his or her research in a particular field of interest. (Reseapro, 2015). For a detailed guide to writing and formatting dissertations for the Izmir Institute of Technology, please consult the following link: http://mfbe.iyte.edu.tr/wp-content/uploads/2014/02/Tez-yaz%C4%B1m-k%C4%B1lavuzu1.pdf

Types of Journal Articles

There are three basic types of journal articles, each with different purposes:

1. Those that present the **results of original research**
2. Those that build on existing research to **offer new interpretations**
3. Those that **review and analyze the current state of knowledge** on a topic (Henderson, 2012).

Original Research Articles

Original research articles publish full reports of data from research. They are also called *Original Articles, Research Articles, Research* or just *Articles,* depending on the journal. They generally include the following sections: Introduction, Methods, Results, and Conclusions/Discussion.

The Master Template of Original Research Article: The general organization and content of academic writing varies according to discipline. However, the master template for an original research article (below) covers articles in many different fields.

__C____ Studies of X indicate...

_____ However, few have investigated...

_____ This study aims to…./ The hypothesis of this study is….

_____ The research approach employed in this study involved…

_____ The results of this study are..

_____ Drawing from the results, one can conclude from this study that...

_____ Result A agrees with the hypothesis/ previous studies. However, result B was surprising because... Weaknesses of this study include... In light of the results and conclusion of this study, room for future research consists of….

Exercise 6: How well do you know the sections of original research articles? Label the above template with the letters (A-G) in the blanks preceding the sentences. The first one has been done for you.

A. Conclusions: What the results may mean
B. Method: The way the study was conducted
C. Literature Review: Summarizing previous research
D. Discussion: Significance of the results and conclusion
E. Research Gap: Explaining what is missing in previous research
F. Thesis: The main purpose or question of the paper
G. Results: What the study found

Sections of Original Research Articles: In-depth descriptions

Title

Academic titles are often lengthy and informative and composed mostly of nouns, many of them specific to the discipline. The title of a scholarly article is designed to give the reader information about the content at a glance. It allows experts and student researchers to understand whether the article will be useful to them (Henderson, 2012).

Abstract

An abstract is a summary of the article. Abstracts precede most journal articles, giving a preview of content by focusing of the study's purpose, method, results, and conclusion. They may also briefly explain the background (for example, the need for the study) or consider the findings' significance. They usually range from 100 to 250 words but can be longer (Henderson, 2012). The abstract is the only section of the article that does not cite information taken from sources. However, the uncited information in the abstract taken from sources is repeated and cited in the introduction.

Introduction / Background

The introduction section, which may or not be labeled as such, prepares the reader for the body of the paper by introducing important concepts or summarizing previous studies on the topic. It usually begins with a review of the relevant literature.

Literature Review

A literature review summarizes related studies on the topic to prepare the way for the unique contribution of the author's own study. The literature review of a journal article summarizes many studies concisely in a short space, often only one to three paragraphs.

Let us examine the second paragraph in a literature review in a paper on brain injury in ice hockey. The first paragraph includes general facts about the prevalence of concussions. In the second paragraph of the introduction, shown below, the authors mention consequences of concussions, especially as they apply to hockey players. As the review continues, it becomes more specific.

> _____ Repeated concussions and TBI [traumatic brain injury] are of particular concern as they may cause life-lasting cognitive and psychosocial deficits [5,6]. _____ These injuries are common in all contact sports, but those who play ice hockey are at particular injury risk [7,8,9,10]. _____ The potential Jong-lasting effects of TBI suggest that these injuries are an important threat to public health [11]. _____ Prevention of sport-related TBIs requires multifaceted approaches that consider issues related to the nature of play and the culture existent within ice hockey [12,13].
> (From Henderson, 2012)

Exercise 7: Label the above four sentences with their purposes (one of the purposes below can be assigned to two of the sentences above):

a. To suggest what a solution to the problem may involve.
b. To explain the relevance of brain injury to the sport of hockey.
c. To explain why brain injury is an important topic.

Justification

Academic authors usually announce how their work will contribute to the field of study. The justification often follows the literature review and answers questions like:

Why is the study important?
How will it advance knowledge about the topic?
What gap will it fill?

The following justification states the gap that the study will fill:

> We have scholarly studies on technical improvements during the nineteenth century, on social interest in the microscope, and on its use by literary figures. But we have little on how the microscopic world itself was perceived . . ..
> - "The Microscopic World" by B. Lightman

The literature review demonstrates the writer's credibility and shows what others have written, while the justification reveals where the author's own study fits in.
(From Henderson, 2012)

Thesis Statement

In scholarly articles, the thesis is near the end of the introduction. In experiments, the thesis may consist of a hypothesis or prediction. The experiment is designed to test the hypothesis, and the conclusion will announce whether it was proven or disproven. Another common thesis form is the "essay plan," a statement of intent outlining the areas to be explored in the order they will appear (Henderson, 2012).

Materials and Methods

This section is also called **Methods, Methodology,** or **Methods and Materials**. In this section you explain how you carried out your study. You describe the object of study, the study location, the experimental or sampling design, the protocol for collecting data, and how the data were analyzed (Bates College, 2011).

Results

Here you objectively present your key results, without interpretation, referring to your figures and tables. You may summarize statistical analyses. The section is organized around tables and or figures. It is written concisely and objectively (Bates College, 2011).

Conclusions and Discussion

In the Conclusion or Discussion section, you interpret your results in light of what was already known about the subject of investigation and explain new understanding of the problem after taking your results into consideration. The Conclusion and Discussion will connect to the Introduction by way of the question(s) or hypotheses and the literature cited. It tells how your study has moved us forward from the place you left us at the end of the Introduction. It also discusses any limitations or weaknesses of the study, and presents areas for further research (Bates College, 2011)

Research Review Articles

Synthesizing sources is particularly important in a research review article. A review article summarizes the research on a particular topic. To write a research review article, the author reads and summarizes existing works on a topic that have already been published in journals and books. Research review articles are good resources to get background information on a topic. They also point out gaps in research and offer important areas for future research. Journal search engines like PubMed, ScienceDirect and Jstor have filter options so you can search exclusively for review articles.

Part II: Close Readings

Reading texts in your field and target genre (journal and magazine articles, dissertations, etc.) is the best way to notice organizational and linguistic patterns that will improve your knowledge, research and writing. The following are close readings with questions that will help you recognize the purpose and style of well-written academic texts.

Close Reading #1: Sample Research Proposal

The following is a research proposal for an article on implementing a needle exchange program (a social service that allows drug users to obtain hypodermic needles at little or no cost) in prisons. Read the proposal and answer the follow-up questions.

Proposal for Research Article on the Effects of Implementing Prison-Based Needle Exchange Programs in Canadian Federal Prisons

by Kate Newcombe

Topic: The benefits that introducing needle exchange programs into the Canadian federal prison system will have on inmates and employees.

Purpose: To investigate prison-based needle exchange programs and argue the benefits of implementing such a system in Canadian federal prisons.

Description: With the recent introduction of the safe injection site in downtown Vancouver, a growing interest in these sites has developed throughout the community, health services programs, and governments. Although it is a controversial topic, evidence from the Vancouver needle exchange site demonstrates the benefits of these programs**. This issue is worth exploring because** drug use continues to be widespread in Canadian prisons, and the increased health risks to intravenous drug users due to lack of proper injecting equipment are growing rapidly. **Currently, no such programs exist** in Canadian prisons. I am interested in discovering more about prison-based needle exchange programs and arguing for the benefits they provide to inmates as well as prison workers. The main organizational methods will be problem-solution and cause-effect.

Tentative Thesis Statement and Central Questions: Prison-based needle exchange programs are an effective, cost-efficient, and beneficial safety tool for public health officials to implement in Canadian prisons in efforts to control drug-related problems and the spread of HIV/AIDS.

• What are the health benefits to intravenous drug users by introducing a system such as this into Canadian prisons?

• Will the introduction of needle exchange systems increase drug use by inmates? • How will its introduction affect prison employees? (i.e., will there be a physical threat to the health and safety of workers?)

• Have other countries implemented this system into their prisons? If so, what are the results? • How, if at all, will the introduction of this system help control the spread of HIV/AIDS in the prison population?

• How has the Canadian government dealt with groups and individuals who argue for implementation?

• Is this truly a cost-effective system?

Methodology: In my preliminary research through my university database, I have found several reliable scholarly articles and reviews of prison-based needle exchange programs. They are peer-reviewed and diverse, from such journals as Addiction, CMAJ, The Lancet, and The New England journal of Medicine. Tentative articles include Dolan, Rutter, and Wodak (2002), "Prison-Based Syringe Exchange Programmes: A Review of International Research and Development", and Bayou mi and Zaric (2008), "The Cost-Effectiveness of Vancouver's Supervised Injection Facility"; other studies available also evaluate the success of the Vancouver program. The researchers' findings support the argument that the introduction of prison-based needle exchange programs is beneficial to inmates and employees, while it does not appear that the health benefits of clean syringes and needles increase intravenous drug use within prisons. Davies' "Prison's Second Death Row" (2004) also looks promising as the author accounts for the reluctance of some governments to institute harm reduction programs.

Follow-up Exercise:

Based on the Sample Research Proposal above, answer the following questions:

1. What is the importance of the topic?

   _____
   _____

2. What is the gap in the research that the author aims to fill?

   _____
   _____

3. Paraphrase the author's main hypothesis that she wishes to test through her study.

   _____
   _____

4. What do you think the author may discover through her research? What conclusions do you predict that she will reach?

   _____
   _____

5. Can you find an instance where the author uses the first person (I or We)? Do you think it is alright to use the first person in academic writing? Why or why not?

   _____
   _____

Close Reading #2: Dissertation Abstract

The following is the 234-word abstract of a doctoral dissertation in Informatics, Computing and Engineering by Timur Gilmanov at Indiana University. To read the full text, visit the following link: https://pqdtopen.proquest.com/doc/2149988180.html?FMT=AI

Lower Bound Resource Requirements for Machine Intelligence

By Tim Gilmanov

Abstract

_____ Recent advancements in technology and the field of artificial intelligence provide a platform for new applications in a wide range of areas, including healthcare, engineering, vision, and natural language processing, that would be considered unattainable one or two decades ago. With the expected compound annual growth rate of 50% during the years of 2017–2021, the field of global artificial intelligence is set to observe increases in computational complexities and amounts of sensor data processed.

_____ In spite of the advancements in the field, truly intelligent machine behavior operating in real time is yet an unachieved milestone. First, in order to quantify such behavior, a definition of machine intelligence would be required, which has not been agreed upon by the community at large. Second, delivering full machine intelligence, as defined in this work, is beyond the scope of today's cutting-edge high-performance computing machines.

_____ One important aspect of machine intelligent systems is resource requirements and the limitations that today's and future machines could impose on such systems. The goal of this research effort is to provide an estimate on the lower bound resource requirements for machine intelligence. A working definition of machine intelligence for purposes of this research is provided, along with definitions of an abstract architecture, workflow, and performance model. Combined together, these tools allow an estimate on resource requirements for problems of machine intelligence, and provide an estimate of such requirements in the future.

Follow-up Exercise 1: Assign the following labels to each paragraph of the above abstract:

A. Background information on recent advancements and the current state of the field of artificial intelligence
B. The goal of the current study: to determine what requirements are needed to obtain true machine intelligence, and how these may be developed in the future
C. The obstacles to realizing authentically intelligent machine behavior/ The limits in the field of artificial intelligence

Follow-up Exercise 2:

1. Which part of this abstract is a mini **literature review**?

   _____

   _____

2. What is the **gap or weakness in the field** that the writer identifies?

   _____
   _____

3. Can you identify the **thesis** or purpose of the dissertation within the abstract?

   _____
   _____

4. How will this dissertation contribute to the field of artificial intelligence? How will it pave the way for future research in the field?

   _____
   _____

5. Why doesn't the author include any **citations** in this text?

   _____
   _____

Close Reading #3: Original Research Article

The following is an article on artificial intelligence applied to cancer detection. It is an original research article that presents new, unique results based on an experiment. It was written by professors of Dermatology and published in the *Annals of Oncology*, a top medical journal on the prevention, diagnosis, and treatment of cancer.

Man against machine: diagnostic performance of deep learning

Authors: H.A. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Blum, A. Kalloo, A. Ben Hadj Hassen, A. Enk & L. Uhlmann

Journal: *Annals of Oncology* 29: 1836-1842, 2018

Pre-reading vocabulary:

- Deep learning: a machine learning method based on artificial neural networks
- Convolutional neural networks (CNN): a type of artificial neural network most commonly applied to analyzing visual imagery
- Melanoma: a type of cancer that typically occurs in the skin
- Dermatology: the branch of medicine dealing with the skin, nails, hair and its diseases
- Dermoscopy: the examination of skin lesions with a tool called a dermatoscope

Close Reading Exercise: Skim this article and notice how it is organized and written, and how the author synthesizes and cites information. You may take notes in the margins. As you read each section, answer the following questions. You do not have to read every detail of the entire article to be able to answer the questions. Skim for relevant information. When you finish, discuss your answers in groups of 3-4.

Abstract (p. 1836)

Background

1. What is the importance of deep learning convolutional neural networks in relation to cancer detection?
   _____
   _____

2. What is the gap in research on CNN in cancer detection mentioned?
   _____
   _____

Conclusions

3. What is the summary of the results?
   _____
   _____

4. What is the main conclusion/recommendation that the authors reach based on the computational results?
   _____
   _____

Introduction (p. 1837)

5. Why is the topic of melanoma important?
   _____
   _____

6. What measures have already been taken to improve detection of melanoma?

_____
_____

7. What are the limitations of current methods of melanoma detection?

_____
_____

8. What revolutionary solution to this problem was proposed in 2017?

_____
_____

9. What is the **thesis** or purpose of the current study? Where in the introduction is the thesis located?

_____
_____

10. Where is the **literature review** located in the Introduction?

_____
_____

11. Does the author employ mostly summary, paraphrase or quotation to cite sources?

_____
_____

Methods (p. 1837)

12. How did the researchers address ethical issues?

_____
_____

13. Which deep learning CNN did the researchers use?

_____
_____

Discussion (p. 1839)

14. Why is it important to improve accuracy of melanoma detection?

_____
_____

15. What did this study do for the first time? That is, how did they contribute in a unique way to the field of cancer detection?

_____
_____

(p. 1841)

16. Name one limitation of the study that the authors recognize.

    _____
    _____

17. How can future studies overcome the limitations that the authors mention?

    _____
    _____

18. Summarize or paraphrase the conclusion that the researchers reach in the last paragraph of their Discussion section.

    _____
    _____

Follow-up questions:

19. Which section(s) include(s) the most summarizing and synthesis of sources? Why?

    _____
    _____

20. Which section(s) was most difficult to understand?

    _____
    _____

21. Which section(s) was easiest to understand?

    _____
    _____

22. Which section(s) would be of most interest to an academician wondering whether or not to employ CNN in melanoma detection?

    _____
    _____

23. Which section(s) would be of most interest to a researcher wishing to repeat a similar study to strengthen results and contribute further to knowledge on this issue?

    _____
    _____

ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

Annals of Oncology 29: 1836–1842, 2018
doi:10.1093/annonc/mdy166
Published online 28 May 2018

ORIGINAL ARTICLE

Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists

H. A. Haenssle[1*,†], C. Fink[1†], R. Schneiderbauer[1], F. Toberer[1], T. Buhl[2], A. Blum[3], A. Kalloo[4], A. Ben Hadj Hassen[5], L. Thomas[6], A. Enk[1] & L. Uhlmann[7]

[1]Department of Dermatology, University of Heidelberg, Heidelberg; [2]Department of Dermatology, University of Göttingen, Göttingen; [3]Office Based Clinic of Dermatology, Konstanz, Germany; [4]Dermatology Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, USA; [5]Faculty of Computer Science and Mathematics, University of Passau, Passau, Germany; [6]Department of Dermatology, Lyons Cancer Research Center, Lyon 1 University, Lyon, France; [7]Institute of Medical Biometry and Informatics, University of Heidelberg, Heidelberg, Germany

*Correspondence to: Prof. Dr med. Holger A. Haenssle, Department of Dermatology, University of Heidelberg, Im Neuenheimer Feld 440, 69120 Heidelberg, Germany. Tel: +49-6221-56-39555; Fax: +49-6221-56-4996; E-mail: Holger.Haenssle@med.uni-heidelberg.de

†Both authors contributed equally as co-first authors.

Background: Deep learning convolutional neural networks (CNN) may facilitate melanoma detection, but data comparing a CNN's diagnostic performance to larger groups of dermatologists are lacking.

Methods: Google's Inception v4 CNN architecture was trained and validated using dermoscopic images and corresponding diagnoses. In a comparative cross-sectional reader study a 100-image test-set was used (level-I: dermoscopy only; level-II: dermoscopy plus clinical information and images). Main outcome measures were sensitivity, specificity and area under the curve (AUC) of receiver operating characteristics (ROC) for diagnostic classification (dichotomous) of lesions by the CNN versus an international group of 58 dermatologists during level-I or -II of the reader study. Secondary end points included the dermatologists' diagnostic performance in their management decisions and differences in the diagnostic performance of dermatologists during level-I and -II of the reader study. Additionally, the CNN's performance was compared with the top-five algorithms of the 2016 International Symposium on Biomedical Imaging (ISBI) challenge.

Results: In level-I dermatologists achieved a mean (±standard deviation) sensitivity and specificity for lesion classification of 86.6% (±9.3%) and 71.3% (±11.2%), respectively. More clinical information (level-II) improved the sensitivity to 88.9% (±9.6%, $P = 0.19$) and specificity to 75.7% (±11.7%, $P < 0.05$). The CNN ROC curve revealed a higher specificity of 82.5% when compared with dermatologists in level-I (71.3%, $P < 0.01$) and level-II (75.7%, $P < 0.01$) at their sensitivities of 86.6% and 88.9%, respectively. The CNN ROC AUC was greater than the mean ROC area of dermatologists (0.86 versus 0.79, $P < 0.01$). The CNN scored results close to the top three algorithms of the ISBI 2016 challenge.

Conclusions: For the first time we compared a CNN's diagnostic performance with a large international group of 58 dermatologists, including 30 experts. Most dermatologists were outperformed by the CNN. Irrespective of any physicians' experience, they may benefit from assistance by a CNN's image classification.

Clinical trial number: This study was registered at the German Clinical Trial Register (DRKS-Study-ID: DRKS00013570; https://www.drks.de/drks_web/).

Key words: melanoma, melanocytic nevi, dermoscopy, deep learning convolutional neural network, computer algorithm, automated melanoma detection

Introduction

Over the past few decades, melanoma has emerged as a major challenge in public health [1]. The continuous increase in incidence rates and melanoma mortality have fueled a heightened commitment to early detection and prevention [2]. Several meta-analyses have shown that dermoscopy significantly improves the diagnostic accuracy of the naked eye examination [3–5]. However, dermatologists and medical practitioners formally trained in different dermoscopic algorithms showed an average sensitivity for detecting melanoma of mostly <80% [6, 7]. In recent years, several strategies of automated computer image analysis have been investigated as an aide for physicians to provide a high and widely reproducible diagnostic accuracy for melanoma screening [8–11]. These approaches were limited by using 'man-made' dermoscopic segmentation criteria for the diagnosis of melanoma (e.g. multiple colors, certain morphological structures as streaks/pseudopods, irregular vascular structures) [12]. As a landmark publication, Esteva et al. reported on the training and testing of a deep learning convolutional neural network (CNN) for imaged-based classification in 2017 [13]. In this setting the CNN was not restricted by man-made segmentation criteria, but deconstructed digital images down to the pixel level and eventually created its own diagnostic clues. As in the study reported herein, the authors utilized a pretrained GoogleNet Inception CNN architecture [14] additionally trained with more than 100 000 digital images and corresponding disease labels.

The aim of the present study was to train, validate, and test a deep learning CNN for the diagnostic classification of dermoscopic images of lesions of melanocytic origin (melanoma, benign nevi) and to compare the results to a large group of 58 dermatologists.

Methods

The study was approved by the local ethics committee and carried out in accordance with the Declaration of Helsinki principles.

Details on methods pertaining to the CNN architecture and CNN training are found in supplementary Methods, available at *Annals of Oncology* online.

We used and specifically trained a modified version of Google's Inception v4 CNN architecture (supplementary Figure S1, available at *Annals of Oncology* online) [14].

Test-set-300

We created a 300-image test-set including 20% melanomas (*in situ* and invasive) of all body sites and of all frequent histotypes, and 80% benign melanocytic nevi of different subtypes and body sites including the so-called 'melanoma simulators' (supplementary Table S1, available at *Annals of Oncology* online). As almost two-third of benign nevi were non-excised lesions validated by follow-up examinations, this dataset represented a spectrum of melanocytic lesions as typically encountered in daily clinical routine. Images of the test-set-300 were retrieved from the high-quality validated image library of the Department of Dermatology, University of Heidelberg, Germany. Various camera/dermoscope combinations were used for image acquisition. No overlap between datasets for training/validation and testing was allowed.

Test-set-100 and reader study level-I and -II

Before CNN testing two experienced dermatologists prospectively selected 100 images of set-300 for an increased diagnostic difficulty (supplementary Table S2, available at *Annals of Oncology* online). Set-100 was used for CNN testing in comparison to dermatologists in a global reader study. Readers ($n = 172$) were invited via mailing lists of the International Dermoscopy Society, and 58 (33.7%) returned their completed voting sheets. Participants indicated their level of experience in dermoscopy ('Beginner' <2 years of experience, 'Skilled' 2–5 years of experience, 'Expert' ≥ 5 years of experience).

In level-I of the reader study, dermatologists were presented solely the dermoscopic image and asked to indicate their dichotomous diagnosis (melanoma, benign nevus) and their management decision (excision, short-term follow-up, send away/no action needed). After an interval of 4 weeks, the same participants indicated their diagnosis and management decision in level-II of the reader study, which included dermoscopic images supplemented by additional clinical information and close-up images of the same 100 cases.

International Symposium on Biomedical Imaging challenge dataset

We used another 100-image dataset created by the International Skin Imaging Collaboration (ISIC) melanoma project for the occasion of the 2016 International Symposium on Biomedical Imaging (ISBI) challenge. This dataset enabled the direct comparison of our CNN to the internationally top-five ranked algorithms [15].

Statistical analysis

The primary outcome measures were sensitivity, specificity, and area under the curve (AUC) of receiver operating characteristics (ROC) for the diagnostic classification (dichotomous) of lesions by the CNN versus dermatologists during level-I or -II of the reader study. Secondary end points included the assessment of the dermatologists' diagnostic performance in their management decisions and the differences in the diagnostic performance of dermatologists between level-I and II of the reader study. For management decisions the option of a 'short-term follow-up' was positively accounted for both sensitivity and specificity calculations. The mean number (percentage) of all lesions and all melanomas indicated for follow-up, the benign nevus excision rate (number of excised nevi/number of all nevi), and the number needed to excise (NNE; number of excised lesions/number of excised melanomas) were calculated.

The CNN put out a 'malignancy score' ranging from 0 to 1 with a cut-off of > 0.5 for the dichotomous classification of malignant versus benign lesions. For comparison of the CNN to dermatologists a two-sided, one-sample *t*-test was applied and the specificity at the level of the average dermatologist sensitivity and the ROC AUC of the CNN versus the mean ROC area of dermatologists was calculated. For dermatologists' dichotomous predictions, area under ROC curves is equivalent to the average of sensitivity and specificity. Descriptive statistics as frequency, mean, range, and standard deviation were used. Two-sided *t*-tests were used to assess differences in the dermatologists' diagnostic performance between level-I and II of the reader study. Results were considered statistically significant at the $P < 0.05$ level. All analyses were carried out using SPSS Version 24 (IBM, SPSS; Chicago, IL).

Results

Dermatologists' diagnostic accuracy

Seventeen (29.3%) out of the 58 participating dermatologists from 17 countries indicated being a 'beginner' in dermoscopy (< 2 years of experience) while 11 (19%) and 30 (51.7%) declared to be 'skilled' (2–5 years of experience) or an 'expert' (> 5 years of

Table 1. Results of reader study level-I and -II

	Classification			Management decision		
Dermatologists	**Sensitivity (%)**	**Specificity (%)**	**ROC area**	**Sensitivity (%)**	**Specificity (%)**	**ROC area**
Level-I						
All (*n*=58)	86.6	71.3	0.79	98.8	64.6	0.82
'Expert' (*n*=30)	89.0	74.5	0.82	98.8	68.1	0.83
'Skilled' (*n*=11)	85.9	68.5	0.77	98.6	61.6	0.80
'Beginner' (*n*=17)	82.9	67.6	0.75	98.8	60.7	0.80
Level-II						
All (*n*=58)	88.9	75.7	0.82	98.6	66.7	0.83
'Expert' (*n*=30)	89.5	77.7	0.84	99.1	69.0	0.84
'Skilled' (*n*=11)	90.9	77.2	0.84	98.2	68.4	0.83
'Beginner' (*n*=17)	86.6	71.2	0.79	98.1	61.3	0.80

ROC, receiver operating characteristic.
Level-I, readers were solely provided with dermoscopic images.
Level-II, readers were additionally provided with clinical information close-up images.
'Expert', the reader indicated to have >5 years of experience in dermoscopy.
'Skilled', the reader indicated to have 2–5 years of experience in dermoscopy.
'Beginner', the reader indicated to have <2 years of experience in dermoscopy.

experience), respectively. Due to reasons of feasibility dermatologists were asked to read only test-set-100.

Diagnostic classification in reader study level-I (dermoscopy only). The mean [±standard deviation (SD)] sensitivity and specificity of the 58 dermatologists for the dichotomous classification of set-100 lesions during study level-I was 86.6% (±9.3%) and 71.3% (±11.2%), respectively (Table 1). This translated into an average (±SD) ROC area of 0.79 (±0.06). Experts in dermoscopy showed a significantly higher mean sensitivity, specificity, and ROC area than beginners [89% (±9.2%), 74.5% (±12.6%), 0.82 (±0.06) versus 82.9% (±7.1%), 67.6% (±6.3%), 0.75 (±0.04), respectively; all $P < 0.02$; Table 1].

Management decisions in reader study level-I (dermoscopy only). Participants were offered (i) excision, (ii) short-term follow-up, or (iii) send away/no action needed as management decisions. In this setting, the average (±SD) sensitivity and ROC area significantly increased to 98.8% (±2.9%, $P < 0.01$) and 0.82 (±0.07, $P = 0.03$), respectively (Table 1). In contrast, the specificity significantly decreased from 71.3% to 64.6% (±13.6%, $P < 0.01$). Similar changes were observed across all levels of experience. Among all dermatologists the average (±SD) benign nevus excision rate was 35.4% (±13.6%) and the lesion follow-up rate was 33.5% (±11.7%). Dermatologists included an average number (±SD) of 1.9 (±1.6) melanomas in follow-up and attained a NNE of 2.3 (±0.6). Higher experience was associated with a significant reduction of the benign nevus excision rate, the lesion follow-up rate, and the number of melanomas under follow-up (all $P < 0.05$). The NNE also slightly improved with experience, however, without reaching statistical significance.

Diagnostic classification in reader study level-II (dermoscopy and clinical information). The addition of clinical information (age, sex, and body site) and close-up images improved the dermatologists' mean (±SD) sensitivity, specificity, and ROC area to 88.9% (±9.6%, $P = 0.19$), 75.7% (±11.7%, $P < 0.05$), and 0.82 (±0.06, $P < 0.01$), respectively (Table 1). These changes were solely based on significant improvements of 'beginners' and 'skilled' dermatologists, while 'experts' in dermoscopy showed no relevant benefit from supplemented clinical information and images.

Management decisions in reader study level-II (dermoscopy and clinical information). When asked for their management decisions during level-II of the study, dermatologists improved their level-II results of the dichotomous classification to a mean (±SD) sensitivity, specificity, and ROC area of 98.6% (±2.8%, $P < 0.01$), 66.7% (±12.4%, $P < 0.01$), and 0.83 (±0.06, $P = 0.76$) (Table 1). However, we found no significant differences between these results and management decision of study level-I. The average (±SD) number of melanomas included into short-term follow-up dropped from 1.9 (±1.6) to 1.3 (±1.5) melanomas ($P = 0.03$) and the NNE remained unchanged at 2.3 benign nevi excised for the detection of one melanoma. For management decisions in study level-II a higher level of experience ('experts' versus 'beginners') was associated with a significantly better mean (±SD) ROC area [0.84 (±0.06) versus 0.79 (±0.06), $P = 0.03$], whereas other parameters of management decisions in study level-II showed no significant differences in relation to the level of experience.

CNN's diagnostic accuracy

Boxplots in Figure 1 show the distribution of melanoma probability scores for benign nevi, *in situ* melanomas, and invasive melanomas. When the aforementioned settings were applied to

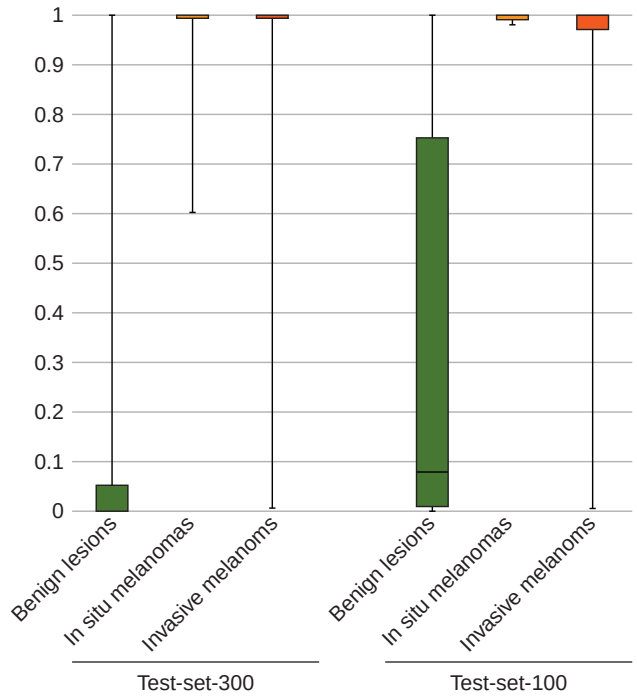

Figure 1. The CNN's melanoma probability scores (range 0–1) for benign nevi (green online) in comparison to *in situ* (orange online) or invasive melanomas (red online) are depicted as boxplots for test-set-300 and test-set-100. Scores closer to 1 indicated a higher probability of melanoma. The upper and lower bounds of boxes indicate the 25th and 75th percentiles while the median is indicated by the line intersection the upper and lower box. Whiskers indicate the full range of probability scores. Statistical analyses revealed significantly different melanoma probability scores when comparing benign lesions to *in situ* or invasive melanomas ($P < 0.001$). However, melanoma probability scores for *in situ* and invasive melanomas showed no significant differences (set-300 $P = 0.84$, set-100 $P = 0.24$).

test-set-100, the sensitivity, specificity, and ROC AUC were 95%, 63.8%, and 0.86, respectively. For the larger test-set-300 including less difficult-to-diagnose lesions the sensitivity, specificity, and ROC AUC were 95%, 80%, and 0.95, respectively. Both ROC curves are depicted in Figure 2A and B.

Diagnostic accuracy of CNN versus dermatologists

We used the dermatologists' mean sensitivity of 86.6% for the diagnostic classification in study level-I as the benchmark for comparison to the CNN (Figure 2A). At this sensitivity the CNN's specificity was higher (82.5%) than the mean specificity of dermatologists (71.3%, $P < 0.01$). Moreover, in level-I the CNN ROC AUC (0.86) was greater than the mean ROC area of dermatologists (0.79, $P < 0.01$).

When dermatologists received more clinical information and images (study level-II) their diagnostic performance improved. Using the dermatologists' level-II mean sensitivity of 88.9% as the operating point on the CNN ROC curve, the CNN specificity was 82.5%, which was significantly higher than the dermatologists' mean specificity of 75.7% ($P < 0.01$). Again, the CNN ROC AUC (0.86) was greater than the mean ROC area of dermatologists (0.82, $P < 0.01$).

CNN comparison to top-five algorithms of ISBI challenge

The head-to-head comparison of ROC curves of our CNN to the international top-five ranked individual algorithms of the ISBI 2016 challenge [15] is shown in Figure 3. With an ROC AUC of 0.79 the CNN presented herein was among the three top algorithms of the ISBI 2016 challenge with almost overlaying ROC curves.

Discussion

Melanoma incidence rates are rising steadily in most fair-skinned populations and were predicted to further increase [2]. Notwithstanding the different levels of training and experience of physicians engaged in early melanoma detection, a reproducible high diagnostic accuracy would be desirable. To this end, we trained and tested a convolutional deep learning CNN for differentiating dermoscopic images of melanoma and benign nevi. For the first time we compared the diagnostic performance of a CNN with a large international group of 58 dermatologists from 17 countries, including 30 experts with more than 5 years of dermoscopic experience. When dermatologists were provided with dermoscopic images only (study level-I) their dichotomous classification of lesions was significantly outperformed by the CNN. However, in a real-life clinical setting dermatologists will incorporate more clinical information into decision-making. Therefore, we investigated the effect of additional clinical information and close-up images and found a much-improved diagnostic performance of dermatologists (study level-II). However, at their improved mean sensitivity (88.9%) dermatologists still showed a specificity inferior to the CNN (75.7% versus 82.5%, $P < 0.01$). Our data clearly show that a CNN algorithm may be a suitable tool to aid physicians in melanoma detection irrespective of their individual level of experience and training. Of note, in study level-I thirteen (22.4%) of 58 dermatologists showed a slightly higher diagnostic performance than the CNN.

We deliberately chose the dermatologists' dichotomous classification of lesions in set-100 as the primary outcome measure for comparison to the CNN. However, it may be argued that 'management decisions' rather than 'diagnostic classifications' represent more the dermatologists' everyday task in skin cancer screenings. Besides 'excision' and 'send away/no action needed' management decisions implied a 'third way', namely the option of a short-term follow-up examination, which was introduced and validated for single lesions with a higher grade of atypia (e.g. variegated tonalities of color, asymmetry in shape, or prominent network) that do not warrant immediate excision for a suspicion of melanoma [16]. The statistical assessment of the follow-up option introduces some difficulties. On the one hand short-term follow-up was shown to be an effective measure to differentiate early melanomas from benign nevi by unmasking dynamic changes [17–19], on the other hand excessive use of the follow-up 'wild-card' (i) may be used to conceal a lack of dermoscopic expertise, (ii) may be largely impracticable in daily clinical routine, and (iii) may delay melanoma excision. Therefore, we positively included the choice to follow-up a lesion into sensitivity (melanomas under follow-up: 'true positives') and specificity calculations (nevi under follow-up: 'true negatives'). However, we also measured details about the use of the

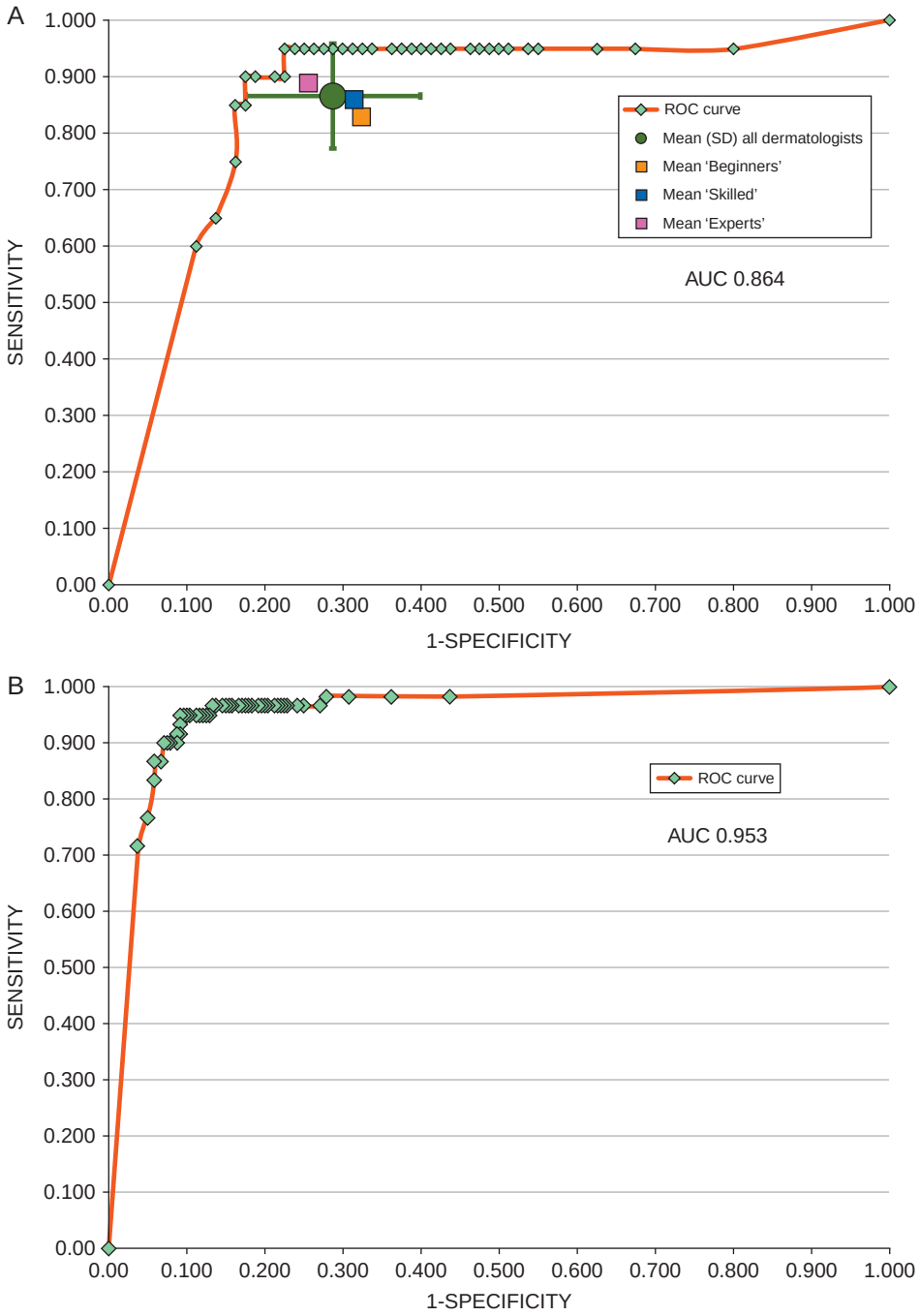

Figure 2. (A) ROC curve of the CNN in relation to the average (±SD) sensitivity and specificity of all dermatologists [mean: green (online) circle; ±SD: green (online) error bars] in set-100 (dichotomous classification, study level-I) and the dermatologists' mean sensitivity and specificity in relation to their level of experience. (B) ROC curve of the CNN in set-300.

follow-up option and found that dermatologists selected approximately one-third of lesions for follow-up, while the mean absolute number of melanomas under follow-up was in the range of 1.3–1.9. As expected, a higher level of experience and more clinical information were associated with reduced follow-up rates.

Important to mention, that differences in the level of difficulty inherent to any image test-set will directly impact the diagnostic performance of algorithms and physicians. In order to generate comparability of different computer algorithms it is therefore of utmost importance to include a large group of dermatologists with various levels of experience as well as to create and use open source datasets as provided by the ISIC [15]. In contrast to Marchetti et al. [15] other authors have not used 'benchmark' image datasets, and only a few studies included a small number of readers for comparison with their designed computer algorithms [13, 20]. Moreover, wherever possible datasets should include lesions of different anatomical sites and histotypes. As shown in supplementary Tables S1 and S2, available at *Annals of Oncology* online, both set-100 and set-300 met these requirements in order to create a less artificial study setting.

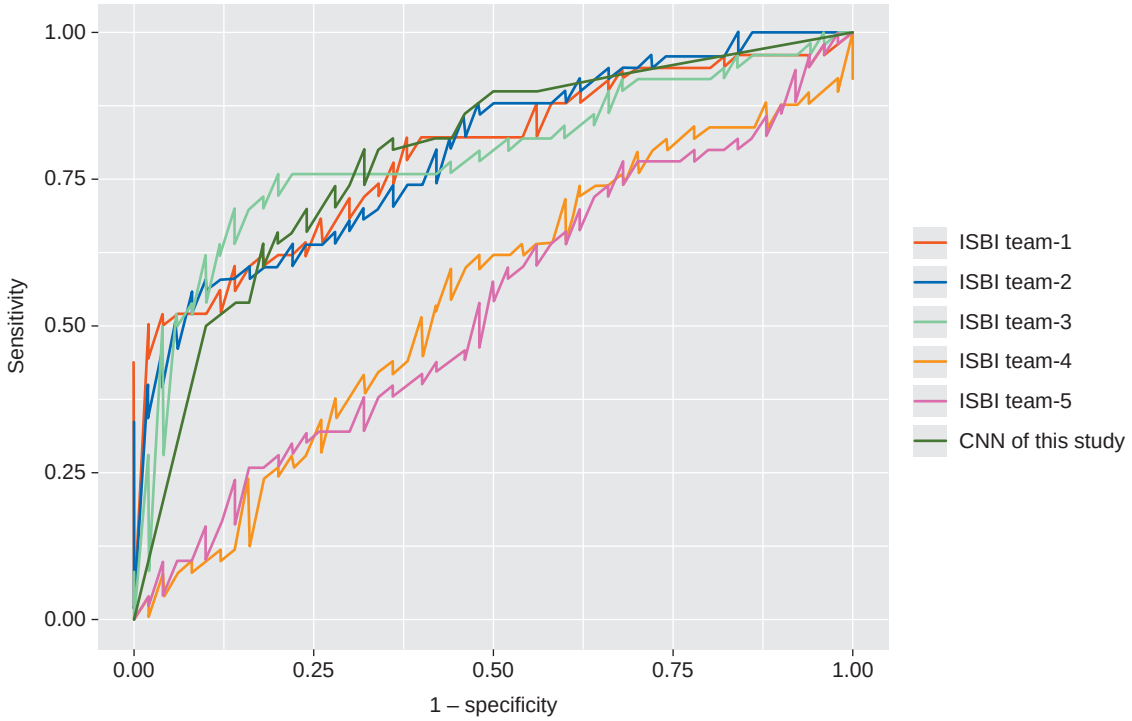

Figure 3. Comparison of ROC curves of the CNN described in this study (dark green (online) line) to the top-five ranked individual algorithms of the 2016 International Symposium on Biomedical Imaging (ISBI) challenge [18]. ROC AUCs in descending order were as follows: ISBI team-2: 0.7956; ISBI team-1: 0.7928; ISBI team-3: 0.7892; CNN of this study: 0.7868; ISBI team-4: 0.5460; ISBI team-5: 0.5324.

Our study shows a number of limitations that may impede a broader generalization. First, as for all reader studies, the setting for testing the dermatologists' diagnostic performance was artificial as they did not need to fear the harsh consequences of missing melanoma. Second, the test-sets of our study did not display the full range of lesions (e.g. pigmented basal cell carcinoma or seborrheic keratosis). Third, the poor availability of validated images led to a shortage of melanocytic lesions from other skin types and genetic backgrounds. Fourth, as shown in earlier reader studies, operating physicians may not follow the recommendations of a CNN they not fully trust, which may diminish the reported diagnostic performance [21]. Besides confirmation of our results with the help of larger and more diverse test-sets, prospective studies are needed that also address the acceptance of patients and physicians involved with screening for skin cancer.

In conclusion, the results of our study demonstrate that an adequately trained deep learning CNN is capable of a highly accurate diagnostic classification of dermoscopic images of melanocytic origin. In conjunction with results from the reader study level-I and -II we could show, that the CNN's diagnostic performance was superior to most but not all dermatologists. While a CNN's architecture is difficult to set up and train, its implementation on digital dermoscopy systems or smart phone applications may easily be deployed. Therefore, physicians of all different levels of training and experience may benefit from assistance by a CNN's image classification.

Acknowledgements

We would like to thank and acknowledge the following dermatologists in alphabetical order, who actively and voluntarily spend much time to participate in the reader study level-I and reader study level-II: Christina Alt, Monika Arenbergerova, Renato Bakos, Anne Baltzer, Ines Bertlich, Andreas Blum, Therezia Bokor-Billmann, Jonathan Bowling, Naira Braghiroli, Ralph Braun, Kristina Buder-Bakhaya, Timo Buhl, Horacio Cabo, Leo Cabrijan, Naciye Cevic, Anna Classen, David Deltgen, Christine Fink, Ivelina Georgieva, Lara-Elena Hakim-Meibodi, Susanne Hanner, Franziska Hartmann, Julia Hartmann, Georg Haus, Elti Hoxha, Raimonds Karls, Hiroshi Koga, Jürgen Kreusch, Aimilios Lallas, Pawel Majenka, Ash Marghoob, Cesare Massone, Lali Mekokishvili, Dominik Mestel, Volker Meyer, Anna Neuberger, Kari Nielsen, Margaret Oliviero, Riccardo Pampena, John Paoli, Erika Pawlik, Barbar Rao, Adriana Rendon, Teresa Russo, Ahmed Sadek, Kinga Samhaber, Roland Schneiderbauer, Anissa Schweizer, Ferdinand Toberer, Lukas Trennheuser, Lyobomira Vlahova, Alexander Wald, Julia Winkler, Priscila Wölbing, Iris Zalaudek. Some participants asked to remain anonymous and we also thank these colleagues for their commitment. Moreover, we thank the International Dermoscopy Society (IDS) for providing the mailing list that enabled the invitation of dermatologists to participate in the study.

Ethical approval

Reviewed and approved by the ethic committee of the medical faculty of the University of Heidelberg (approval number S-629/2017).

Funding

This research received no specific grant from any public, commercial or not-for-profit sector.

Disclosure

HAH received honoraria and/or travel expenses from companies involved in the development of devices for skin cancer screening: Scibase AB, FotoFinder Systems GmbH, Heine Optotechnik GmbH, Magnosco GmbH. CF received travel expenses from Magnosco GmbH. The remaining authors declared no conflicts of interest.

References

1. Koh HK. Melanoma screening: focusing the public health journey. Arch Dermatol 2007; 143(1): 101–103.
2. Nikolaou V, Stratigos AJ. Emerging trends in the epidemiology of melanoma. Br J Dermatol 2014; 170(1): 11–19.
3. Vestergaard ME, Macaskill P, Holt PE, Menzies SW. Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting. Br J Dermatol 2008; 159: 669–676.
4. Bafounta ML, Beauchet A, Aegerter P, Saiag P. Is dermoscopy (epiluminescence microscopy) useful for the diagnosis of melanoma? Results of a meta-analysis using techniques adapted to the evaluation of diagnostic tests. Arch Dermatol 2001; 137(10): 1343–1350.
5. Salerni G, Teran T, Puig S et al. Meta-analysis of digital dermoscopy follow-up of melanocytic skin lesions: a study on behalf of the International Dermoscopy Society. J Eur Acad Dermatol Venereol 2013; 27(7): 805–814.
6. Dolianitis C, Kelly J, Wolfe R, Simpson P. Comparative performance of 4 dermoscopic algorithms by nonexperts for the diagnosis of melanocytic lesions. Arch Dermatol 2005; 141(8): 1008–1014.
7. Carli P, Quercioli E, Sestini S et al. Pattern analysis, not simplified algorithms, is the most reliable method for teaching dermoscopy for melanoma diagnosis to residents in dermatology. Br J Dermatol 2003; 148(5): 981–984.
8. Barata C, Celebi ME, Marques JS. Improving dermoscopy image classification using color constancy. IEEE J Biomed Health Inform 2015; 19: 1–52.
9. Glaister J, Wong A, Clausi DA. Segmentation of skin lesions from digital images using joint statistical texture distinctiveness. IEEE Trans Biomed Eng 2014; 61(4): 1220–1230.
10. Garnavi R, Aldeen M, Bailey J. Computer-aided diagnosis of melanoma using border and wavelet-based texture analysis. IEEE Trans Inform Technol Biomed 2012; 16(6): 1239–1252.
11. Kaya S, Bayraktar M, Kockara S et al. Abrupt skin lesion border cutoff measurement for malignancy detection in dermoscopy images. BMC Bioinformatics 2016; 17(S13): 367.
12. Pehamberger H, Steiner A, Wolff K. In vivo epiluminescence microscopy of pigmented skin lesions. I. Pattern analysis of pigmented skin lesions. J Am Acad Dermatol 1987; 17(4): 571–583.
13. Esteva A, Kuprel B, Novoa RA et al. Dermatologist-level classification of skin cancer with deep neural networks. Nature 2017; 542(7639): 115–118.
14. Szegedy C, Vanhoucke V, Ioffe S et al. Rethinking the inception architecture for computer vision 2015. https://arxiv.org/abs/1512.00567 (9 May 2018, date last accessed).
15. Marchetti MA, Codella NCF, Dusza SW et al. Results of the 2016 International Skin Imaging Collaboration International Symposium on Biomedical Imaging challenge: comparison of the accuracy of computer algorithms to dermatologists for the diagnosis of melanoma from dermoscopic images. J Am Acad Dermatol 2018; 78(2): 270–277.
16. Menzies SW, Gutenev A, Avramidis M et al. Short-term digital surface microscopic monitoring of atypical or changing melanocytic lesions. Arch Dermatol 2001; 137(12): 1583–1589.
17. Altamura D, Avramidis M, Menzies SW. Assessment of the optimal interval for and sensitivity of short-term sequential digital dermoscopy monitoring for the diagnosis of melanoma. Arch Dermatol 2008; 144(4): 502–506.
18. Menzies SW, Emery J, Staples M et al. Impact of dermoscopy and short-term sequential digital dermoscopy imaging for the management of pigmented lesions in primary care: a sequential intervention trial. Br J Dermatol 2009; 161(6): 1270–1277.
19. Menzies SW, Stevenson ML, Altamura D, Byth K. Variables predicting change in benign melanocytic nevi undergoing short-term dermoscopic imaging. Arch Dermatol 2011; 147(6): 655–659.
20. Ferris LK, Harkes JA, Gilbert B et al. Computer-aided classification of melanocytic lesions using dermoscopic images. J Am Acad Dermatol 2015; 73(5): 769–776.
21. Hauschild A, Chen SC, Weichenthal M et al. To excise or not: impact of MelaFind on German dermatologists' decisions to biopsy atypical lesions. J Dtsch Dermatol Ges 2014; 12(7): 606–614.

Close Reading #4: Original Research Article

Changes in Diet and Lifestyle and Long-Term Weight Gain in Women and Men

Authors: D. Mozaffarian, M.D., Dr. P.H. Tao Hao, M.P.H., Eric B. Rimm, Sc.D., Walter C. Willett, MD, Dr. P.H., and Frank B. Hu, M.D., Ph.D.

Journal: *The New England Journal of Medicine* 364: 2392-404, 2011

The following is an article on the effect of lifestyle behavior changes on weight gain. It is an original research article that presents new, unique results based on a longitudinal study. It was written by professors of doctors of Harvard Medical School and published in *The New England Journal of Medicine,* a top medical journal.

Close Reading Exercise: Skim this article and notice how it is organized and written, and how the author synthesizes and cites information. You may take notes in the margins. As you read each section, answer the following questions. You do not have to read every detail of the entire article to be able to answer the questions. Skim for relevant information. When you finish, discuss your answers in groups of 3-4.

Introduction (p.2393)

1. Why is the prevention of weight gain an important topic worth studying?
   _____
   _____
2. What is the gap in the literature? What are the limitations of previous weight-loss trials?
   _____
   _____
3. What are some lifestyle behaviors that may influence people's weight over the long term, according to the literature?
   _____
   _____
4. Do the cited studies 4-10 agree or disagree with each other? What is the conclusion of these studies?
   _____
   _____
5. Do the cited studies 11-14 agree or disagree with each other?
   _____
   _____
6. What is the main question that this study hopes to answer?
   _____
   _____
7. What does the study do in order to answer this question?
   _____
   _____

Methods (p. 2393-2395)

8. Who were the participants involved and what was the timeline of the study?

   _____
   _____

9. Which lifestyle habits were studied?

   _____
   _____

10. How were weight changes assessed?

    _____
    _____

11. How was the relationship between lifestyle changes and weight changes assessed?

    _____
    _____

Figure 1 (p. 2398)

12. Introduce and describe Figure 1 in your own words for someone who has not read the article.

    _____
    _____

Discussion (2398)

13. What were the principal findings of the research?

    _____
    _____

14. How do the findings relate to previous research? (For instance, which data agrees with previous research, and which disagrees?)

    _____
    _____

15. What are the limitations of the study that the authors recognize?

    _____
    _____

16. Summarize the implications and recommendations derived from the data.

    _____
    _____

ORIGINAL ARTICLE

Changes in Diet and Lifestyle and Long-Term Weight Gain in Women and Men

Dariush Mozaffarian, M.D., Dr.P.H., Tao Hao, M.P.H., Eric B. Rimm, Sc.D., Walter C. Willett, M.D., Dr.P.H., and Frank B. Hu, M.D., Ph.D.

ABSTRACT

BACKGROUND

Specific dietary and other lifestyle behaviors may affect the success of the straightforward-sounding strategy "eat less and exercise more" for preventing long-term weight gain.

METHODS

We performed prospective investigations involving three separate cohorts that included 120,877 U.S. women and men who were free of chronic diseases and not obese at baseline, with follow-up periods from 1986 to 2006, 1991 to 2003, and 1986 to 2006. The relationships between changes in lifestyle factors and weight change were evaluated at 4-year intervals, with multivariable adjustments made for age, baseline body-mass index for each period, and all lifestyle factors simultaneously. Cohort-specific and sex-specific results were similar and were pooled with the use of an inverse-variance–weighted meta-analysis.

RESULTS

Within each 4-year period, participants gained an average of 3.35 lb (5th to 95th percentile, −4.1 to 12.4). On the basis of increased daily servings of individual dietary components, 4-year weight change was most strongly associated with the intake of potato chips (1.69 lb), potatoes (1.28 lb), sugar-sweetened beverages (1.00 lb), unprocessed red meats (0.95 lb), and processed meats (0.93 lb) and was inversely associated with the intake of vegetables (−0.22 lb), whole grains (−0.37 lb), fruits (−0.49 lb), nuts (−0.57 lb), and yogurt (−0.82 lb) ($P \leq 0.005$ for each comparison). Aggregate dietary changes were associated with substantial differences in weight change (3.93 lb across quintiles of dietary change). Other lifestyle factors were also independently associated with weight change ($P<0.001$), including physical activity (−1.76 lb across quintiles); alcohol use (0.41 lb per drink per day), smoking (new quitters, 5.17 lb; former smokers, 0.14 lb), sleep (more weight gain with <6 or >8 hours of sleep), and television watching (0.31 lb per hour per day).

CONCLUSIONS

Specific dietary and lifestyle factors are independently associated with long-term weight gain, with a substantial aggregate effect and implications for strategies to prevent obesity. (Funded by the National Institutes of Health and others.)

From the Division of Cardiovascular Medicine (D.M.) and Channing Laboratory (D.M., E.B.R., W.C.W., F.B.H.), Brigham and Women's Hospital and Harvard Medical School; and the Departments of Epidemiology (D.M., T.H., E.B.R., W.C.W., F.B.H.) and Nutrition (D.M., E.B.R., W.C.W., F.B.H.), Harvard School of Public Health — all in Boston. Address reprint requests to Dr. Mozaffarian at 665 Huntington Ave., Bldg. 2-319, Boston, MA 02115, or at dmozaffa@hsph.harvard.edu.

N Engl J Med 2011;364:2392-404.

Because efforts to lose weight pose tremendous challenges, primary prevention of weight gain is a global priority. Since weight stability requires a balance between calories consumed and calories expended, the advice to "eat less and exercise more" would seem to be straightforward. However, weight gain often occurs gradually over decades (about 1 lb per year), making it difficult for most people to perceive the specific causes. Weight-loss trials[1-3] have typically enrolled obese or overweight persons who attempted substantial short-term weight loss on specialized diets, thus limiting the generalizability of the findings to nonobese populations and to the factors that determine long-term, gradual weight gain.

Several lifestyle behaviors may influence whether or not a person can maintain energy balance over the long term. For instance, the consumption of sugar-sweetened beverages, sweets, and processed foods may make it harder to do so, whereas the consumption of whole grains, fruits, and vegetables might make it easier.[4-10] Physical activity should also influence long-term weight gain, but evidence to support this expectation has been surprisingly inconsistent.[11-14] In addition, the duration of television viewing and of sleep may influence energy consumption, energy expenditure, or both.[15-19] Different lifestyle behaviors have often been evaluated separately, thus limiting relative comparisons or the quantification of combined effects. In addition, most studies of long-term weight gain have evaluated current behaviors, but changes in behavior over time may be more relevant in terms of both their biologic effects on long-term weight gain and their translation into prevention strategies. We investigated the relationship between multiple lifestyle changes, both independently and jointly, and long-term weight gain in nonobese women and men participating in three separate, prospective studies.

METHODS

STUDY DESIGN AND POPULATION

The Nurses' Health Study (NHS) is a prospective study of a cohort of 121,701 female registered nurses from 11 U.S. states who were enrolled in 1976. The Nurses' Health Study II (NHS II) is a prospective study of a cohort of 116,686 younger female registered nurses from 14 states who were enrolled in 1989. The Health Professionals Follow-up Study (HPFS) is a prospective study of a cohort of 51,529 male health professionals from all 50 states, enrolled in 1986. Participants were followed with the use of biennial validated questionnaires concerning medical history, lifestyle, and health practices. For this analysis, the baseline year was the first year for which detailed information was available on diet, physical activity, and smoking habits — 1986 in the NHS and HPFS and 1991 in the NHS II. We excluded participants with obesity, diabetes, cancer, or cardiovascular, pulmonary, renal, or liver disease at baseline; those for whom baseline data on lifestyle habits were missing; those with an implausible energy intake (<900 or >3500 kcal per day); those with more than nine blank responses on the diet questionnaire; those who were newly pregnant during follow-up; and those who were over 65 years of age, given possible confounding due to age-related loss of lean muscle mass. The final analyses included 50,422 women in the NHS, 47,898 women in the NHS II, and 22,557 men in the HPFS, all of whom were free of obesity and chronic diseases and for whom data on weight and lifestyle habits at baseline were complete. Cohort members who were excluded because of missing data had characteristics similar to those included in the analysis (data not shown). The funders of this study had no role in its design or conduct; in the collection, management, analysis, or interpretation of the data; or in the preparation, review, or approval of the manuscript.

LIFESTYLE ASSESSMENT

Lifestyle habits of interest were physical activity, television watching, alcohol use, sleep duration, and diet, and cigarette smoking was a potential confounding factor (Table 1, and Tables 1 and 2 in the Supplementary Appendix, available with the full text of this article at NEJM.org). On the basis of their plausible biologic effects, the dietary factors we assessed included fruits, vegetables, whole grains, refined grains, potatoes (including boiled or mashed potatoes and french fries), potato chips, whole-fat dairy products, low-fat dairy products, sugar-sweetened beverages, sweets and desserts, processed meats, unprocessed red meats, fried foods, and trans fat (see Table 1 in the Supplementary Appendix). We also evaluated nuts, 100%-fruit juices, diet sodas, and subtypes of dairy products and potatoes. Different types of alcohol drinks were also evaluated. To assess aggregate dietary

Table 1. Baseline Characteristics and Average 4-Year Lifestyle Changes among 120,877 U.S. Women and Men in Three Prospective Cohorts.*

Variable	Nurses' Health Study (N=50,422 women)		Nurses' Health Study II (N=47,898 women)		Health Professionals Follow-up Study (N=22,557 men)	
	Baseline	Change within Each 4-Year Period (5th–95th percentile)†	Baseline	Change within Each 4-Year Period (5th–95th percentile)†	Baseline	Change within Each 4-Year Period (5th–95th percentile)†
Age — yr	52.2±7.2		37.5±4.1		50.8±7.5	
Weight — lb	141±20	2.33 (–5.5 to 10.7)‡	138±20	5.24 (–2.3 to 16.3)‡	175±20	1.63 (–5.0 to 8.0)‡
Body-mass index	23.7±1.4	0.40 (–0.94 to 1.83)	23.0±2.7	0.69 (–1.05 to 2.82)	24.7±1.1	0.24 (–0.70 to 1.20)
Physical activity — MET-hr/wk§	14.8±9.9	1.78 (–9.24 to 15.3)	21.6±25.9	0.53 (–12.6 to 14.7)	22.9±15.1	6.7 (–8.90 to 33.6)
Alcohol — drinks/day	0.52±0.41	–0.02 (–0.35 to 0.22)	0.30±0.52	0.04 (–0.14 to 0.33)	0.87±0.60	0.00 (–0.45 to 0.43)
Total daily sleep — hr¶	7.0±0.5	—	7.0±1.0	—	7.1±0.5	—
Dietary intake — servings/day						
Fruits	1.6±0.6	–0.10 (–0.96 to 0.59)	1.2±0.9	0.03 (–0.48 to 0.57)	1.5±0.6	0.05 (–0.57 to 0.72)
Vegetables	3.9±1.0	–0.45 (–2.32 to 0.78)	3.3±1.9	0.19 (–0.84 to 1.41)	3.3±0.9	0.02 (–1.15 to 1.07)
Whole-fat dairy foods	1.2±0.5	–0.16 (–0.86 to 0.36)	0.8±0.7	–0.02 (–0.45 to 0.50)	0.9±0.5	–0.09 (–0.64 to 0.36)
Low-fat dairy foods	0.9±0.5	0.11 (–0.46 to 0.83)	1.1±0.9	0.02 (–0.59 to 0.64)	0.8±0.5	–0.05 (–0.64 to 0.43)
Potato chips	0.1±0.1	–0.01 (–0.09 to 0.07)	0.2±0.2	–0.01 (–0.12 to 0.10)	0.2±0.1	–0.01 (–0.11 to 0.09)
Potatoes	0.4±0.1	–0.01 (–0.19 to 0.18)	0.4±0.3	–0.02 (–0.17 to 0.12)	0.4±0.1	–0.01 (–0.18 to 0.18)
Whole grains	0.5±0.2	0.12 (–0.22 to 0.53)	0.7±0.5	0.10 (–0.23 to 0.49)	0.7±0.3	0.10 (–0.23 to 0.52)
Refined grains	1.2±0.4	–0.03 (–0.60 to 0.57)	1.3±0.8	–0.09 (–0.67 to 0.40)	1.2±0.5	–0.01 (–0.57 to 0.56)
Nuts	0.2±0.1	0.01 (–0.15 to 0.20)	0.1±0.1	0.06 (–0.05 to 0.33)	0.2±0.2	0.02 (–0.22 to 0.29)
Sugar-sweetened beverages	0.2±0.2	–0.00 (–0.14 to 0.11)	0.3±0.6	–0.03 (–0.31 to 0.14)	0.3±0.3	–0.02 (–0.22 to 0.14)
Diet soda, 0 calories	0.5±0.4	–0.01 (–0.40 to 0.38)	1.0±1.3	–0.06 (–0.73 to 0.52)	0.5±0.5	0.00 (–0.35 to 0.36)
100%-fruit juice	0.8±0.4	0.06 (–0.43 to 0.64)	0.6±0.7	–0.03 (–0.45 to 0.33)	0.8±0.4	0.01 (–0.44 to 0.50)
Sweets and desserts	1.3±0.6	–0.01 (–0.75 to 0.75)	1.3±1.1	–0.06 (–0.65 to 0.45)	1.5±0.7	–0.03 (–0.76 to 0.70)
Processed meats	0.3±0.1	–0.04 (–0.22 to 0.11)	0.2±0.2	0.01 (–0.14 to 0.19)	0.4±0.2	–0.02 (–0.22 to 0.17)
Unprocessed red meats	0.6±0.2	–0.06 (–0.32 to 0.14)	0.6±0.4	–0.01 (–0.24 to 0.21)	0.6±0.2	0.02 (–0.28 to 0.41)
Time spent watching television — hr/wk‖	4.2±0.7	—	8.4±7.5	0.25 (–5.00 to 6.08)	10.5±4.0	–0.29 (–5.50 to 5.00)

* Plus–minus values are means ±SD. Data are based on 20 years of follow-up (1986–2006) in the Nurses' Health Study (NHS), 12 years of follow-up (1991–2003) in the Nurses' Health Study II (NHS II), and 20 years of follow-up (1986–2006) in the Health Professionals Follow-up Study (HPFS). Usual dietary habits and alcohol use were assessed every 4 years with the use of validated, semiquantitative food-frequency questionnaires. Alcohol included wine (5 oz), beer (1 glass, bottle, or can), and liquor (1 drink or shot). Other specific foods and beverages in each category are listed in Table 1 in the Supplementary Appendix. Results for smoking habits (assessed with the use of biennial questionnaires), trans fat, and fried foods can be found in Table 2 in the Supplementary Appendix. To convert pounds to kilograms, divide by 0.45.

† Because serial assessments were limited, the 4-year changes could not be reliably quantified, and absolute levels at baseline were used for cohort-specific analyses.

‡ The corresponding values for relative weight changes were 1.83% (5th to 95th percentile, –4.0 to 7.7) in the NHS, 3.60% (5th to 95th percentile, –1.80 to 10.7) in the NHS II, and 1.00% (5th to 95th percentile, –3.11 to 5.09) in the HPFS.

§ Physical activity was assessed by means of validated questionnaires every 2 years, with average energy expenditure (metabolic-equivalent [MET]-hours per week) for specific activities (e.g., walking, jogging, bicycling, swimming, racquet sports, and gardening). In the NHS II, physical activity levels in 1997, 2001, and 2005 were used to impute the levels in 1995, 1999, and 2003, respectively.

¶ The average duration of sleep per 24-hour period was assessed in 1986, 2000, and 2002 in the NHS; in 2001 in the NHS II; and in 1987 and 2000 in the HPFS.

‖ The average number of hours per week spent watching television at home was assessed in 1992 and 2004 in the NHS; in 1991, 1997, and 2001 in the NHS II; and in 1998 and every 2 years thereafter in the HPFS.

effects, changes in each dietary factor independently associated with weight gain were categorized in quintiles and assigned ascending values (1 to 5) or descending values (5 to 1) for habits inversely or positively associated with weight gain, respectively; these ordinal values were summed to generate an overall score for dietary change.

WEIGHT CHANGES

Height and weight were assessed by questionnaire at enrollment, and weight was requested on each follow-up questionnaire. In a validation subsample, questionnaire-reported and staff-measured weights were highly correlated ($r=0.96$; mean difference, 3.3 lb). Weight changes were evaluated every 4 years as both absolute changes (pounds) and relative changes (percentages). Anthropometric measurements and weight changes strongly predict disease outcomes in these cohorts.[20-26]

STATISTICAL ANALYSIS

We assessed the independent relationships between changes in lifestyle behaviors and weight changes within 4-year periods over a period of 20 years in the NHS, 12 years in the NHS II, and 20 years in the HPFS (Table 1), using multivariable linear regression with robust variance and accounting for within-individual repeated measures. Lifestyle changes were assessed either as continuous variables, with censoring of data at the 0.5 and 99.5 percentiles to minimize the influence of outliers, or as indicator variables for categorical behaviors (e.g., smoking status). Potential nonlinear effects of decreases versus increases in each behavior were evaluated by modeling changes in indicator categories, with "no change" as the reference. Any missing lifestyle data during any follow-up period were coded as a missing indicator category for categorical variables (e.g., smoking status) and with carried-forward values for continuous variables. To minimize confounding from the loss of lean muscle mass at older ages or from loss of weight due to undiagnosed chronic disease, we censored data for participants after they reached 65 years of age or if they received a diagnosis of chronic pulmonary, renal, or liver disease or of cancer other than nonmelanoma skin cancer (3 to 4% of participants); data were censored 6 years before diagnosis to account for preclinical disease. Multivariable models were used to adjust for age, baseline body-mass index in each 4-year period, and all lifestyle factors simultaneously. Total energy intake, biologic factors (e.g., blood pressure), and medications were not included as covariables because such factors could be mediators (in causal pathways) or direct correlates of mediators of the effects of lifestyle on weight gain. Sensitivity analyses were performed to evaluate absolute physical activity levels at the start of each 4-year period rather than changes during the period, data among participants who never smoked, and data stratified according to age and baseline body-mass index. Findings across cohorts were pooled by means of inverse-variance–weighted, random-effects meta-analyses. Analyses were carried out with the use of SAS software, version 9.1 (SAS Institute), at a two-tailed alpha level of 0.05.

RESULTS

BASELINE CHARACTERISTICS AND WEIGHT GAIN

Weight and lifestyle characteristics at baseline and changes during 1,570,808 person-years of follow-up are shown in Table 1. The mean weight gain for all the 4-year periods combined differed among the three cohorts, a finding that may have been related to cohort-specific differences in sex and age at baseline. The mean weight gains were as follows: 1.63 lb (5th to 95th percentile, −5.0 to 8.0) for the men in the HPFS (mean age, 50.8±7.5 years), 2.33 lb (5th to 95th percentile, −5.5 to 10.7) for the women in the NHS (mean age, 52.2±7.2), and 5.24 lb (5th to 95th percentile, −2.3 to 16.3) for the women in the NHS II (mean age, 37.5±4.1). The average weight gain across the cohorts was 3.35 lb (5th to 95th percentile, −4.1 to 12.4), or 2.4% of body weight (5th to 95th percentile, −3.0 to 8.4), during each 4-year period; this change corresponds to a weight gain of 16.8 lb over a period of 20 years.

DIET AND LIFESTYLE CHANGES

Although the mean changes in lifestyle in the overall study population were small, the between-individual changes were large (Table 1). In the NHS, for example, the difference between persons in the upper level of change and those in the lower level of change (95th percentile minus 5th percentile) was 3.1 servings per day for vegetable consumption, 25.3 metabolic equivalents (METs) per week for physical activity, and 0.66 drinks per day for alcohol consumption. Correlations between various lifestyle changes were generally small ($r<0.05$). Positive correlations in changes were

largest for fruits and vegetables (r=0.21) and for processed meats and unprocessed red meats (r=0.21), whereas inverse correlations in changes were largest for whole-fat dairy products and low-fat dairy products (r=−0.08).

RELATIONSHIPS BETWEEN DIETARY CHANGES AND WEIGHT CHANGES

After multivariable adjustment, nearly every dietary factor was independently related to weight change (Table 2). Findings were similar, in direction and magnitude, for men and women and across the three cohorts (Fig. 1). (For additional detailed sex-specific and cohort-specific results, see Tables 3, 4, and 5 in the Supplementary Appendix.) The dietary factors with the largest positive associations with weight changes, per serving per day, were increases in the consumption of potato chips (1.69 lb), potatoes (1.28 lb), sugar-sweetened beverages (1.00 lb), unprocessed red meats (0.95 lb), and processed meats (0.93 lb). A secondary analysis of potato subtypes showed that weight changes were positively associated with increases in the consumption of french fries (3.35 lb) and of boiled, baked, or mashed potatoes (0.57 lb). Weight gain associated with increased consumption of refined grains (0.39 lb per serving per day) was similar to that for sweets and desserts (0.41 lb per serving per day). Inverse associations with weight gain, per serving per day, were seen for increased consumption of vegetables (−0.22 lb), whole grains (−0.37 lb), fruits (−0.49 lb), nuts (−0.57 lb), and yogurt (−0.82 lb).

Categorical analyses of each dietary factor revealed similar linear relationships for increased versus decreased consumption (data not shown) — that is, for each dietary factor, the weight change with increased consumption was the inverse of that with decreased consumption. Thus, less weight gain occurred with decreased consumption of potato chips, processed meats, sugar-sweetened beverages, potatoes, or trans fat, and more weight gain occurred with decreased consumption of vegetables, whole grains, fruits, nuts, or yogurt. Aggregate dietary changes were robustly related to weight gain in a dose-dependent fashion, with a 3.93-lb greater weight gain across quintiles (Table 3).

RELATIONSHIPS BETWEEN OTHER LIFESTYLE FACTORS AND WEIGHT CHANGES

Other lifestyle behaviors were also independently related to weight change (Table 3), with similar findings in men and women and across the three cohorts. (For sex-specific and cohort-specific results, see Tables 6, 7, and 8 in the Supplementary Appendix.) Across quintiles, participants with greater increases in physical activity gained 1.76 fewer pounds within each 4-year period. Absolute levels of physical activity, rather than changes in these levels, were not associated with weight change (data not shown). Overall, increases in alcohol use (per drink per day) were positively associated with weight change (0.41 lb), but heterogeneity was evident with respect to both the beverage type and the size and direction of changes in use (see the figure in the Supplementary Appendix). Sleep duration had a U-shaped association with weight gain, with greater weight gain occurring with less than 6 hours or more than 8 hours of sleep per night. Increases in time spent watching television (per hour per day) were independently associated with weight gain (0.31 lb, P<0.001).

As compared with persons who never smoked, those who had quit smoking within the previous 4 years had a weight gain of 5.17 lb (Table 3). Subsequent weight gain for former smokers was small (0.14 lb per 4-year period). Continued smoking was inversely associated with weight gain (−0.70 lb), a finding that may have been related to undiagnosed chronic disease. Initiation of smoking was not associated with weight change, but evaluation of this category was limited by its rarity in these populations (accounting for <0.1% of person-years). Findings for other lifestyle factors were similar when restricted to persons who never smoked (Table 3).

ADDITIONAL ANALYSES

We categorized dietary changes more finely to examine wider ranges of potential effects (Fig. 2). As compared with participants in the top decile, participants in each lower decile of dietary change had greater weight gain, in a dose-dependent fashion. Changes in diet and physical activity were independently associated with weight gain (Fig. 2). Findings for all lifestyle factors were generally similar in analyses stratified according to age or baseline body-mass index (the weight in kilograms divided by the square of the height in meters) (<25 or 25 to 30), although the magnitude of associated weight change was generally larger among overweight persons (Tables 9 and 10 in the Supplementary Appendix). All results were also similar when we evaluated relative (percent) weight changes rather than absolute weight changes (not shown).

Table 2. Pooled, Multivariable-Adjusted Results for the Relationships between Changes in Dietary Habits and Weight Change.*

Increased Dietary Intake†	Weight Change within Each 4-Year Period (95% Confidence Interval)†			
	Age-Adjusted Change *lb*	P Value	Multivariable-Adjusted Change‡ *lb*	P Value
Fruits	−0.69 (−0.92 to −0.46)	<0.001	−0.49 (−0.63 to −0.35)	<0.001
Vegetables	−0.25 (−0.37 to −0.12)	<0.001	−0.22 (−0.34 to −0.11)	<0.001
Nuts	−0.78 (−1.31 to −0.26)	<0.001	−0.57 (−0.97 to −0.17)	0.005
Whole-fat dairy foods§	0.25 (0.05 to 0.45)	0.01	0.10 (−0.10 to 0.30)	0.31
Butter	0.47 (0.23 to 0.71)	<0.001	0.30 (0.15 to 0.45)	<0.001
Cheese	0.13 (−0.08 to 0.34)	0.23	0.02 (−0.09 to 0.13)	0.75
Whole-fat milk	0.08 (−0.05 to 0.22)	0.24	−0.06 (−0.19 to 0.06)	0.32
Low-fat dairy foods§	−0.17 (−0.21 to −0.13)	<0.001	−0.05 (−0.14 to 0.05)	0.33
Low-fat or skim milk	−0.02 (−0.11 to 0.07)	0.69	0.06 (−0.07 to 0.20)	0.37
Yogurt¶	−1.16 (−1.48 to −0.84)	<0.001	−0.82 (−0.99 to −0.67)	<0.001
Potato chips	3.01 (2.09 to 3.94)	<0.001	1.69 (1.30 to 2.09)	<0.001
Potatoes§	2.14 (1.26 to 3.03)	<0.001	1.28 (0.87 to 1.70)	<0.001
French fried	6.59 (4.35 to 8.83)	<0.001	3.35 (2.29 to 4.42)	<0.001
Boiled, baked, or mashed	0.99 (0.36 to 1.61)	0.002	0.57 (0.26 to 0.89)	<0.001
Whole grains‖	−0.59 (−0.65 to −0.53)	<0.001	−0.37 (−0.48 to −0.25)	<0.001
Refined grains	0.56 (0.28 to 0.83)	<0.001	0.39 (0.21 to 0.58)	<0.001
Sugar-sweetened beverages	1.32 (1.03 to 1.62)	<0.001	1.00 (0.83 to 1.17)	<0.001
100%-fruit juice	0.19 (0.07 to 0.31)	0.002	0.31 (0.14 to 0.47)	<0.001
Diet soda	−0.12 (−0.23 to −0.02)	0.02	−0.11 (−0.21 to −0.02)	0.02
Sweets or desserts	0.65 (0.31 to 1.00)	<0.001	0.41 (0.16 to 0.66)	0.001
Processed meats	1.76 (1.43 to 2.09)	<0.001	0.93 (0.79 to 1.08)	<0.001
Unprocessed red meats	1.68 (1.08 to 2.28)	<0.001	0.95 (0.55 to 1.34)	<0.001
Trans fat	1.44 (0.95 to 1.94)	<0.001	0.65 (0.41 to 0.89)	<0.001
Fried foods				
Consumed at home	0.49 (0.33 to 0.64)	<0.001	0.36 (0.22 to 0.51)	<0.001
Consumed away from home	0.52 (0.27 to 0.76)	<0.001	0.28 (0.08 to 0.48)	0.007

* Data are based on 20 years of follow-up (1986–2006) in the Nurses Health Study, 12 years of follow-up (1991–2003) in the Nurses Health Study II, and a 20 years of follow-up (1986–2006) in the Health Professionals Follow-up Study. Findings according to sex and within each study were generally similar to the pooled results (see the tables in the Supplementary Appendix). To convert pounds to kilograms, divide by 0.45.

† The weight changes shown are for increased consumption; decreased consumption would be associated with the inverse of these weight changes. Increased consumption was defined as an increase in the number of servings per day for all items except trans fat (an increase in the percent of energy) and fried foods consumed at home or away from home (an increase in the number of servings per week).

‡ Values were adjusted for age, baseline body-mass index at the beginning of each 4-year period, and sleep duration, as well as for changes in physical activity, alcohol use, television watching, smoking, and all the dietary factors in the table simultaneously.

§ For the categories of whole-fat dairy foods, low-fat dairy foods, and potatoes, subtypes were evaluated together in the full, multivariable-adjusted model in place of the overall food group (e.g., butter, cheese, and whole-fat milk were evaluated in place of total whole-fat dairy foods).

¶ We were unable to evaluate separately the different types of yogurt (e.g., nonfat, low-fat, or whole-fat; sweetened or unsweetened). U.S. consumption patterns would suggest that most participants chose nonfat or low-fat yogurt; however, no inference could be made with regard to sweetened or unsweetened yogurt.

‖ Findings were similar when either total dietary fiber or cereal fiber was included in the analysis instead of whole grains.

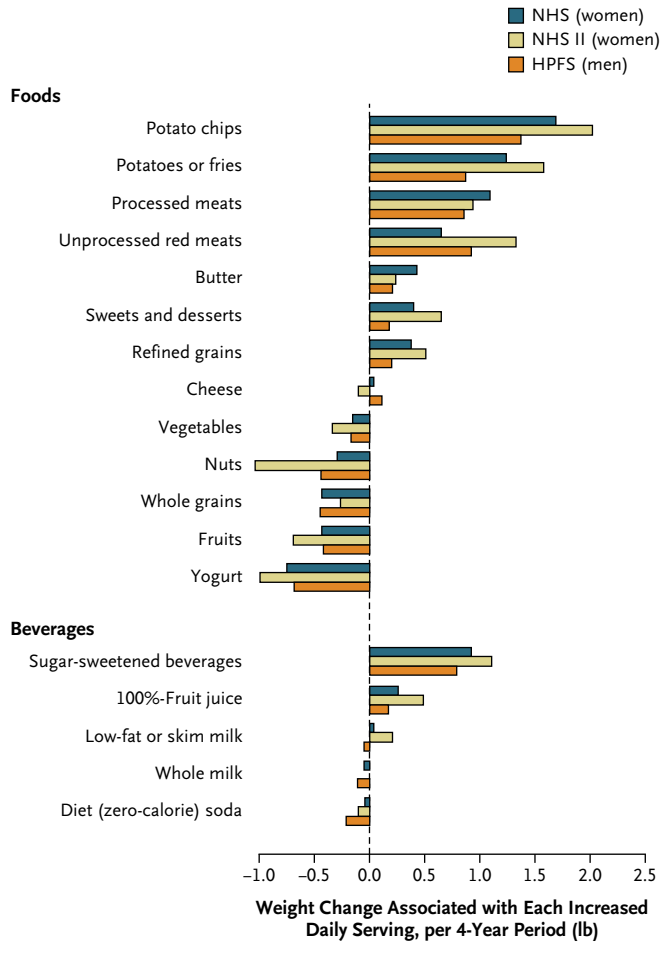

Figure 1. Relationships between Changes in Food and Beverage Consumption and Weight Changes Every 4 Years, According to Study Cohort.

Study participants included 50,422 women in the Nurses' Health Study (NHS), followed for 20 years (1986 to 2006); 47,898 women in the Nurses' Health Study II (NHS II), followed for 12 years (1991 to 2003); and 22,557 men in the Health Professionals Follow-up Study (HPFS), followed for 20 years (1986 to 2006). Weight changes are reported for each increase in the daily serving of the food or beverage; decreased intake would be associated with the inverse weight changes. There was little evidence of a significant interaction between diet and physical activity ($P>0.10$ for the interaction in each cohort). All weight changes were adjusted simultaneously for age, baseline body-mass index, sleep duration, and changes in smoking status, physical activity, television watching, alcohol use, and all of the dietary factors shown. The P value is less than 0.001 for all dietary factors with the exception of butter in the NHS II, cheese in the NHS and NHS II, low-fat or skim milk in the NHS and HPFS, diet soda in the NHS, and whole-fat milk in all three cohorts.

DISCUSSION

We found that multiple lifestyle changes were independently associated with long-term weight gain, including changes in the consumption of specific foods and beverages, physical activity, alcohol use, television watching, and smoking habits. Average long-term weight gain in nonobese populations is gradual — in the cohorts we studied, about 0.8 lb per year — but accumulated over time, even modest increases in weight have implications for long-term adiposity-related metabolic dysfunction, diabetes, cardiovascular disease, and cancer.[21-24] Whereas weight changes associated with any single lifestyle factor were relatively modest in our three cohorts, in the aggregate, changes in diet and physical activity accounted for large differences in weight gain. The results were similar across the three separate cohorts, increasing our confidence in the validity and generalizability of the findings.

All these relationships must be mediated by changes in energy intake, energy expenditure, or both. Total energy intake is not well estimated from dietary questionnaires, nor does it reflect energy balance, which is necessarily codetermined by energy expenditure. Thus, weight change is the best population metric of energy imbalance and at least partly captures energy intake after adjustment for determinants of expenditure (e.g., age, body-mass index, and physical activity).

Eating more or less of any one food or beverage may change the total amount of energy consumed, but the magnitude of associated weight gain varied for specific foods and beverages. Differences in weight gain seen for specific foods and beverages could relate to varying portion sizes, patterns of eating, effects on satiety, or displacement of other foods or beverages. Strong positive associations with weight change were seen for starches, refined grains, and processed foods. These findings are consistent with those suggested by the results in limited short-term trials: consumption of starches and refined grains may be less satiating, increasing subsequent hunger signals and total caloric intake, as compared with equivalent numbers of calories obtained from less processed, higher-fiber foods that also contain healthy fats and protein.[27] Consumption of processed foods that are higher in starches, refined grains, fats, and sugars can increase weight gain.[28-30]

Some foods — vegetables, nuts, fruits, and whole grains — were associated with less weight gain when consumption was actually increased. Obviously, such foods provide calories and cannot violate thermodynamic laws. Their inverse associations with weight gain suggest that the increase in their consumption reduced the intake of other foods to a greater (caloric) extent, decreasing the

Table 3. Pooled, Multivariable-Adjusted Results for the Relationships between Changes in Lifestyle Habits and Weight Change.*

Lifestyle Habit	Weight Change Every 4 Years (95% Confidence Interval)			
	Age-Adjusted Change *lb*	P Value	Multivariable-Adjusted Change† *lb*	P Value
Dietary change, median score‡				
Quintile 1: 42	3.97 (2.85 to 5.10)	<0.001	3.93 (2.88 to 4.98)	<0.001
Quintile 2: 47	2.32 (1.99 to 2.66)	<0.001	2.41 (1.96 to 2.87)	<0.001
Quintile 3: 50	2.08 (1.35 to 2.80)	<0.001	2.10 (1.37 to 2.83)	<0.001
Quintile 4: 54	1.64 (1.22 to 2.06)	<0.001	1.64 (1.22 to 2.05)	<0.001
Quintile 5: 60	Reference		Reference	
Physical-activity change, median change in MET-hr/wk				
Quintile 1: –16.3	Reference		Reference	
Quintile 2: –2.59	–0.07 (–0.36 to 0.22)	0.64	–0.07 (–0.32 to 0.18)	0.60
Quintile 3: 1.59	–0.85 (–1.11 to –0.59)	<0.001	–0.81 (–1.02 to –0.60)	<0.001
Quintile 4: 6.49	–1.00 (–1.10 to –0.91)	<0.001	–0.92 (–1.02 to –0.83)	<0.001
Quintile 5: 23.2	–1.86 (–2.31 to –1.41)	<0.001	–1.76 (–2.14 to –1.38)	<0.001
Increase in no. of alcohol drinks/day§	0.39 (0.23 to 0.55)	<0.001	0.41 (0.23 to 0.59)	<0.001
Total daily hr of sleep¶				
<6	Reference		Reference	
6 to 7	–0.36 (–0.43 to –0.29)	<0.001	–0.32 (–0.38 to –0.25)	<0.001
>7 to 8	–0.35 (–0.48 to –0.22)	<0.001	–0.30 (–0.42 to –0.18)	<0.001
>8	–0.09 (–1.11 to 0.92)	0.86	–0.09 (–0.96 to 0.78)	0.84
Missing data	–0.49 (–0.93 to –0.06)	0.03	–0.41 (–0.72 to –0.10)	0.01
Increase in hr/day watching television‖	0.32 (0.13 to 0.51)	0.001	0.31 (0.20 to 0.42)	<0.001
Change in smoking status				
Never smoked, no change	Reference		Reference	
Former smoker, no change	0.19 (0.13 to 0.25)	<0.001	0.14 (0.08 to 0.20)	<0.001
Current smoker, no change	–0.62 (–0.97 to –0.26)	<0.001	–0.70 (–1.10 to –0.31)	<0.001
Current smoker, changed to former smoker	4.99 (3.89 to 6.09)	<0.001	5.17 (4.06 to 6.29)	<0.001
Former smoker, changed to current smoker	–2.47 (–3.82 to –1.12)	<0.001	–2.81 (–4.24 to –1.38)	<0.001
Never smoked, changed to current smoker	0.67 (–1.48 to 2.82)	0.02	0.32 (–2.31 to 2.95)	0.81

* Data are based on a 20-year follow-up (1986–2006) in the Nurses' Health Study (NHS), a 12-year follow-up (1991–2003) in the Nurses' Health Study II (NHS II), and a 20-year follow-up (1986–2006) in the Health Professionals Follow-up Study (HPFS). Findings within each study and restricted to persons who never smoked were generally similar to these pooled results (see the tables in the Supplementary Appendix). To convert pounds to kilograms, divide by 0.45.

† Values were adjusted for age, baseline body-mass index at the beginning of each 4-year period, and all the variables shown in the table simultaneously.

‡ Changes in dietary habits associated with weight change were derived by totaling the ordinal values for the quintiles of change for each dietary habit in ascending order (1 to 5) or descending order (5 to 1) for habits that were inversely or positively associated with weight gain, respectively. Scores ranged from 17 to 85.

§ Heterogeneity was evident across beverage types and changes (see the Supplementary Appendix for details).

¶ Data are based on absolute levels, owing to limited data on serial assessments of sleep for each cohort to assess change in each 4-year period.

‖ Differences in weight change are for increased television watching; the inverse difference would be associated with decreased watching. Values are based on data from the NHS II and the HPFS; limited serial assessments precluded an analysis of changes in television watching in the NHS.

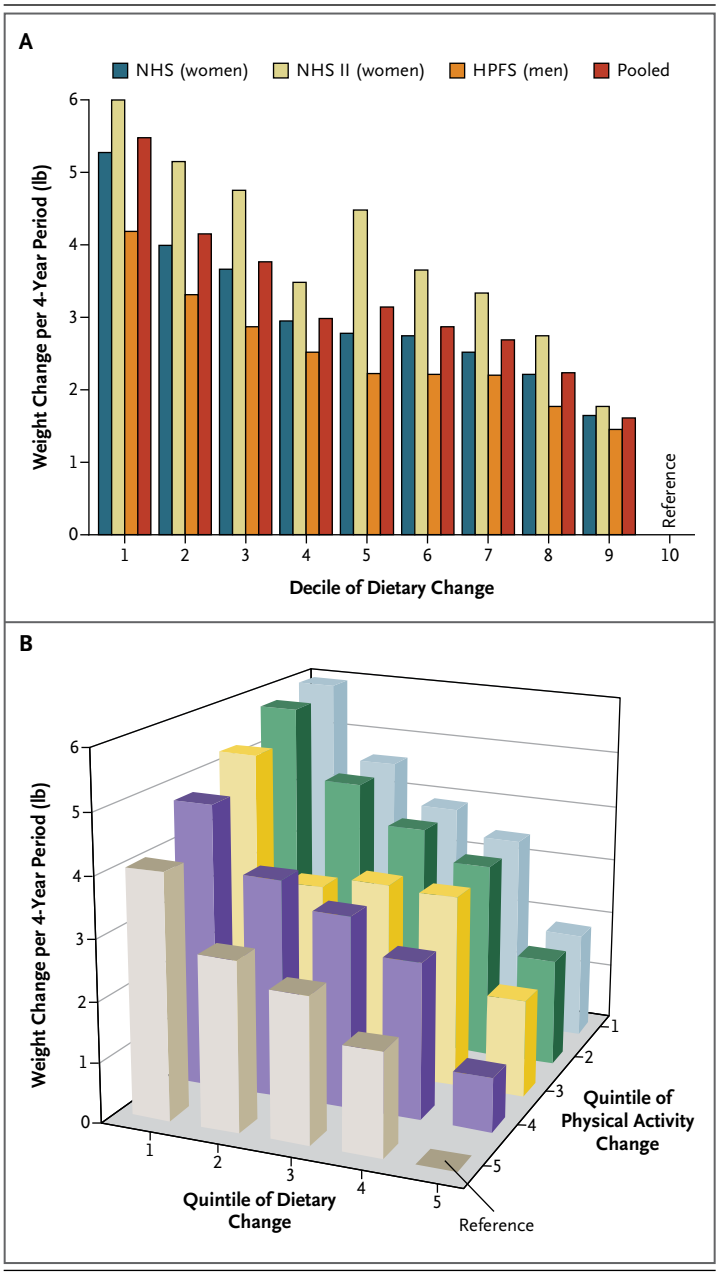

Figure 2. Relationships between Changes in Diet and Physical Activity and Weight Changes within Each 4-Year Period in the Three Cohorts.

In a multivariable-adjusted analysis, overall dietary changes among the 120,877 men and women in the three cohorts were based on the sum of changes in the intake of fruits, vegetables, whole grains, nuts, refined grains, potatoes or french fries, potato chips, butter, yogurt, sugar-sweetened beverages, 100%-fruit juice, sweets and desserts, processed meats, unprocessed red meats, trans fat, fried foods consumed at home, and fried foods consumed away from home. Panel A shows the relationship between deciles of dietary change and weight change per 4-year period in the three cohorts separately and combined. As compared with persons in the top decile, persons in the bottom decile had a 5.48-lb greater weight gain (95% confidence interval [CI], 4.02 to 6.94). Panel B shows the relationship between the cross-stratified quintiles of changes in both dietary habits and physical activity with weight changes per 4-year period for the combined cohorts. As compared with persons in the top quintiles of both dietary change and physical-activity change, persons in the lowest quintiles had a 5.93-lb greater weight gain (95% CI, 4.35 to 7.52). There was little evidence of a significant interaction between diet and physical activity ($P>0.10$ for the interaction in each cohort). All weight changes were adjusted for age, baseline body-mass index, sleep duration, and changes in smoking status, physical activity, television watching, and alcohol use. $P<0.001$ for all comparisons.

overall amount of energy consumed. Higher fiber content and slower digestion of these foods would augment satiety, and their increased consumption would also displace other, more highly processed foods in the diet, providing plausible biologic mechanisms whereby persons who eat more fruits, nuts, vegetables, and whole grains would gain less weight over time.

Yogurt consumption was also associated with less weight gain in all three cohorts. Potential mechanisms for these findings are unclear; intriguing evidence suggests that changes in colonic bacteria might influence weight gain.[31] It is also possible that there is an unmeasured confounding factor that tracks with yogurt consumption (e.g., people who change their yogurt consumption may have other weight-influencing behaviors that were not measured by our instruments).

Our findings with regard to sugar-sweetened beverages are consistent with the results of prior observational studies and short-term interventions.[7,32,33] Consumption of 100%-fruit juice was associated with weight gains of smaller magnitude, possibly because these beverages may be consumed in smaller servings than are sugar-sweetened beverages or in different patterns (i.e., single rather than multiple servings).[33] Findings have been inconsistent in prior studies of alcohol use and weight gain.[34-37] In a previous analysis of alcohol consumption in relation to weight change in the NHS II cohort over a period of 8 years, the smallest weight gain was seen among women who remained moderate drinkers.[36] The present findings suggest that the relationship between alcohol use and weight change is com-

plex, and further analyses are needed that address potential heterogeneity with respect to sex, beverage type, baseline intake, direction of change, and duration of follow-up. Short-term controlled trials suggest that liquids are less satiating than solid foods, increasing the total amount of energy consumed.[38] Overall, our analysis showed that changes in the consumption of all liquids except milk were positively associated with weight gain; our findings for high-carbohydrate beverages were consistent with those for refined carbohydrates and starches consumed in foods. Temporal trends render our findings especially relevant: between 1965 and 2002, U.S. beverage consumption increased from 11.8 to 21.0% of all calories consumed — 222 more kilocalories per person per day — with sugar-sweetened beverages and alcohol accounting for 60% and 32% of the increase, respectively.[39]

Our analysis showed relatively neutral associations between change in the consumption of most dairy foods and weight gains. Few prior studies have evaluated these relationships. Prior analyses of HPFS data showed associations similar to ours for the overall categories of whole-fat and low-fat dairy products,[40] but subtypes (e.g., milk, cheese, and butter) were not evaluated independently. Among Swedish women, higher intakes of whole milk and cheese were inversely associated with weight gain; as in our study, significant associations with weight gain were not seen for other dairy foods.[41] In several long-term studies, inverse associations between dairy consumption and the risk of insulin resistance, the metabolic syndrome, or diabetes were observed,[42,43] but potential mediating effects on weight change were not evaluated. Limited short-term studies of dairy foods and satiety or weight change have had inconsistent results.[44,45]

Overall, our analysis showed divergent relationships between specific foods or beverages and long-term weight gain, suggesting that dietary quality (the types of foods and beverages consumed) influences dietary quantity (total calories). Several dietary metrics that are currently emphasized, such as fat content, energy density, and added sugars, would not have reliably identified the dietary factors that we found to be associated with long-term weight gain. For example, most of the foods that were positively associated with weight gain were starches or refined carbohydrates; no significant differences were seen for low-fat and skim milk versus whole-fat milk, and the consumption of nuts was inversely associated with weight gain. Clear patterns were also not seen in the relationship between weight change and the energy density of dietary components (e.g., beverages of low energy density were strongly associated with weight gain). Foods that contained higher amounts of refined carbohydrates — whether these were added (e.g., in sweets and desserts) or were not added (e.g., in refined grains) — were associated with weight gain in similar ways, and potato products (which are low in sugars and high in starches) showed the strongest associations with weight gain. No single metric appears to capture these complexities. Our findings highlight gaps in our mechanistic understanding of how particular dietary characteristics alter energy balance, suggesting directions for future research regarding pathways involved in hunger, satiety, absorption, metabolism, and adipocyte growth or hyperplasia. In general, changes in the consumption of refined or processed foods and liquid carbohydrates or alcohol were positively associated with weight gain, whereas changes in the consumption of unprocessed foods such as whole grains, fruits, nuts, and vegetables were inversely associated with weight gain. These results suggest that future policies and research efforts to prevent obesity should consider food structure and processing as potentially relevant dietary metrics.

Changes in physical activity were independently related to long-term changes in weight, supporting the biologic plausibility of our overall findings. Prior, smaller studies have shown inverse associations between activity changes and weight change.[11,13] Prevalent (current) levels of physical activity are inconsistently related to weight change, with associations observed only for subgroups of persons[14] or subtypes of activities.[12] As seen in prior analyses of sugar-sweetened beverages,[33] changes in lifestyle may be most relevant for weight gain. Persons may achieve a new steady-state weight within months after a change in regular physical activity, diet, or other lifestyle habits, highlighting the importance of repeated assessments of over time to discern long-term effects.

Many prior studies of television watching and obesity have been cross-sectional, limiting the ability to make inferences about which came first.[15] In controlled interventions, decreased television watching reduced weight gain in children,[16,17] an effect that was mediated more by improvements in dietary habits than by a change in physical activity.

Television watching appears to encourage snacking during viewing and also influences food choices both during viewing and at other times.[46-52] Our long-term prospective evaluation provides evidence that both the duration of television viewing and changes in the duration of viewing influence weight gain in adults. Because these effects are probably mediated by changes in diet and physical activity, and may also be mediated by changes in sleep, the multivariable (mediator)–adjusted associations may underestimate the full effects of television watching.

Decreases in sleep duration are concordant with the U.S. obesity epidemic.[18,19,53] Data from cross-sectional studies and some prospective studies, including a prior analysis of NHS data, support the relationship of shorter sleep duration with obesity.[18] In short-term trials, reduced sleep alters leptin, ghrelin, subjective hunger, and preferences for calorie-dense, refined-carbohydrate foods.[19] Our results suggest that the association between sleep duration and long-term weight gain is characterized by a U-shaped curve — that is, weight gain is lowest among persons who sleep 6 to 8 hours a night and is higher among those who sleep less than 6 hours or more than 8 hours. Future studies should evaluate how changes in sleep over time are related to weight gain.

Our long-term follow-up data confirm prior observations that smoking cessation results in weight gain initially but in little weight change thereafter. The health benefits of cessation exceed any potential adverse effects — that is, active smokers are at higher risk for cardiovascular diseases, cancer, and diabetes than are former smokers.[54] Smoking may also adversely alter the distribution of body fat, promoting visceral rather than femoral or subcutaneous fat deposition; thus, even in the setting of lower total weight, active smoking has adverse metabolic consequences, as evidenced, for example, by its links to a higher risk of type 2 diabetes.[55] Any relative weight loss seen with active smoking should not be considered beneficial, nor should the relative weight gain soon after smoking cessation be considered harmful.

Our study has some limitations. Although dietary questionnaires specified portion sizes, residual, unmeasured differences in portion sizes among participants might account for additional independent effects on energy balance. For example, an average, large baked potato contains 278 calories, as compared with 500 to 600 calories for a large serving of french fries.[56] The typical portion size of a specific food or beverage may therefore partly mediate its effects on weight gain (i.e., both average portion sizes and biologic effects). As for lifestyle behaviors, each was measured with some degree of error, which, if random, would underestimate their true relationships with weight change. Lifestyle changes were self-selected, and residual confounding from other lifestyle behaviors is possible. However, in contrast to prevalent behaviors, changes in these behaviors were generally not strongly correlated ($r<|0.05|$), which suggests that different behaviors are often changed relatively independently, thus minimizing potential confounding. A person's weight change could lead to changes in lifestyle rather than vice versa. Such reverse causation would generally underestimate true effects. For example, persons who are gaining weight might plausibly either reduce their intake of sugar-sweetened beverages and sweets or increase their consumption of vegetables, leading to reverse bias with respect to the observed associations.

As is the case with any biologic finding or medical intervention, our results represent the average population effect, and intraindividual variations exist. The cohorts studied here largely comprised white, educated U.S. adults, which potentially limits the generalizability of the findings. Conversely, the ranges of dietary intakes were broad and overlapped with national estimates. In addition, our findings were broadly consistent with cross-sectional national trends with respect to diet and obesity: between 1971 and 2004, the average dietary intake of calories in the United States increased by 22% among women and by 10% among men, primarily owing to the increased consumption of refined carbohydrates, starches, and sugar-sweetened beverages.[39] Our findings were also consistent among the three cohorts and in analyses stratified according to smoking status, age, and baseline body-mass index, and it seems plausible that the biologic effects of many lifestyle factors would be qualitatively similar in other populations.

A habitual energy imbalance of about 50 to 100 kcal per day may be sufficient to cause the gradual weight gain seen in most persons.[57,58] This means that unintended weight gain occurs easily but also that modest, sustained changes in lifestyle could mitigate or reverse such an energy imbalance. Our findings suggest that both individual and population-based strategies to help people consume fewer calories may be most effec-

tive when particular foods and beverages are targeted for decreased (or increased) consumption. Aggregate dietary changes accounted for substantial differences in weight gain, with additional contributions from changes in physical activity and television watching, thus highlighting specific lifestyle changes that might be prioritized in obesity-prevention strategies.

Supported by grants (DK46200, DK58845, HL085710, HL60712, HL35464, CA87969, CA50385, CA55075, and CA95589) from the National Institutes of Health and by the Searle Scholars Program.

Dr. Mozaffarian reports receiving consulting fees from Nutrition Impact and Foodminds, lecture fees from Aramark, Unilever, and SPRIM, royalties from UpToDate, and grant support on behalf of Harvard Medical School from GlaxoSmithKline, Sigma-Tau, and Pronova and being listed as a coinventor on a provisional patent application filed by and assigned to Harvard University for the use of trans-palmitoleic acid to prevent and treat insulin resistance, type 2 diabetes, and related conditions. Dr. Hu reports receiving lecture fees from Amgen, Nutrition Impact, Unilever, and the Institute of Food Technologies and grant support on behalf of the Harvard School of Public Health from Merck and the California Walnut Commission. No other potential conflict of interest relevant to this article was reported.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank Dr. Donna Spiegelman for statistical advice, Dr. Francine Laden for comments during the preparation of an early version of the manuscript, and Dr. Peilin Shi for assistance with statistical programming.

REFERENCES

1. Gardner CD, Kiazand A, Alhassan S, et al. Comparison of the Atkins, Zone, Ornish, and LEARN diets for change in weight and related risk factors among overweight premenopausal women: the A TO Z Weight Loss Study: a randomized trial. JAMA 2007; 297:969-77. [Erratum, JAMA 2007;298:178.]
2. Shai I, Schwarzfuchs D, Henkin Y, et al. Weight loss with a low-carbohydrate, Mediterranean, or low-fat diet. N Engl J Med 2008;359:229-41. [Erratum, N Engl J Med 2009;361:2681.]
3. Sacks FM, Bray GA, Carey VJ, et al. Comparison of weight-loss diets with different compositions of fat, protein, and carbohydrates. N Engl J Med 2009;360:859-73.
4. Ello-Martin JA, Ledikwe JH, Rolls BJ. The influence of food portion size and energy density on energy intake: implications for weight management. Am J Clin Nutr 2005;82:Suppl 1:236S-241S.
5. van Dam RM, Seidell JC. Carbohydrate intake and obesity. Eur J Clin Nutr 2007; 61:Suppl 1:S75-S99.
6. Rosenheck R. Fast food consumption and increased caloric intake: a systematic review of a trajectory towards weight gain and obesity risk. Obes Rev 2008;9:535-47.
7. Wolff E, Dansinger ML. Soft drinks and weight gain: how strong is the link? Medscape J Med 2008;10:189.
8. Micha R, Mozaffarian D. Trans fatty acids: effects on metabolic syndrome, heart disease and diabetes. Nat Rev Endocrinol 2009;5:335-44.
9. Liu S, Willett WC, Manson JE, Hu FB, Rosner B, Colditz G. Relation between changes in intakes of dietary fiber and grain products and changes in weight and development of obesity among middle-aged women. Am J Clin Nutr 2003;78:920-7.
10. Buijsse B, Feskens EJ, Schulze MB, et al. Fruit and vegetable intakes and subsequent changes in body weight in European populations: results from the project on Diet, Obesity, and Genes (DiOGenes). Am J Clin Nutr 2009;90:202-9.
11. Haapanen N, Miilunpalo S, Pasanen M, Oja P, Vuori I. Association between leisure time physical activity and 10-year body mass change among working-aged men and women. Int J Obes Relat Metab Disord 1997;21:288-96.
12. Wagner A, Simon C, Ducimetière P, et al. Leisure-time physical activity and regular walking or cycling to work are associated with adiposity and 5 y weight gain in middle-aged men: the PRIME Study. Int J Obes Relat Metab Disord 2001;25:940-8.
13. Gordon-Larsen P, Hou N, Sidney S, et al. Fifteen-year longitudinal trends in walking patterns and their impact on weight change. Am J Clin Nutr 2009;89:19-26.
14. Lee IM, Djoussé L, Sesso HD, Wang L, Buring JE. Physical activity and weight gain prevention. JAMA 2010;303:1173-9.
15. Robinson TN. Television viewing and childhood obesity. Pediatr Clin North Am 2001;48:1017-25.
16. Robinson TN. Reducing children's television viewing to prevent obesity: a randomized controlled trial. JAMA 1999;282: 1561-7.
17. Epstein LH, Roemmich JN, Robinson JL, et al. A randomized trial of the effects of reducing television viewing and computer use on body mass index in young children. Arch Pediatr Adolesc Med 2008; 162:239-45.
18. Patel SR, Hu FB. Short sleep duration and weight gain: a systematic review. Obesity (Silver Spring) 2008;16:643-53.
19. Van Cauter E, Spiegel K, Tasali E, Leproult R. Metabolic consequences of sleep and sleep loss. Sleep Med 2008;9:Suppl 1: S23-S28.
20. Chan JM, Rimm EB, Colditz GA, Stampfer MJ, Willett WC. Obesity, fat distribution, and weight gain as risk factors for clinical diabetes in men. Diabetes Care 1994;17:961-9.
21. Willett WC, Manson JE, Stampfer MJ, et al. Weight, weight change, and coronary heart disease in women: risk within the 'normal' weight range. JAMA 1995; 273:461-5.
22. Colditz GA, Willett WC, Rotnitzky A, Manson JE. Weight gain as a risk factor for clinical diabetes mellitus in women. Ann Intern Med 1995;122:481-6.
23. Rexrode KM, Hennekens CH, Willett WC, et al. A prospective study of body mass index, weight change, and risk of stroke in women. JAMA 1997;277:1539-45.
24. Eliassen AH, Colditz GA, Rosner B, Willett WC, Hankinson SE. Adult weight change and risk of postmenopausal breast cancer. JAMA 2006;296:193-201.
25. Thygesen LC, Grønbaek M, Johansen C, Fuchs CS, Willett WC, Giovannucci E. Prospective weight change and colon cancer risk in male US health professionals. Int J Cancer 2008;123:1160-5.
26. Flint AJ, Hu FB, Glynn RJ, et al. Excess weight and the risk of incident coronary heart disease among men and women. Obesity (Silver Spring) 2010;18:377-83.
27. Bornet FR, Jardy-Gennetier AE, Jacquet N, Stowell J. Glycaemic response to foods: impact on satiety and long-term weight regulation. Appetite 2007;49:535-53.
28. Rolls BJ, Roe LS, Beach AM, Kris-Etherton PM. Provision of foods differing in energy density affects long-term weight loss. Obes Res 2005;13:1052-60.
29. Ello-Martin JA, Roe LS, Ledikwe JH, Beach AM, Rolls BJ. Dietary energy density in the treatment of obesity: a year-long trial comparing 2 weight-loss diets. Am J Clin Nutr 2007;85:1465-77.
30. Ledikwe JH, Rolls BJ, Smiciklas-Wright H, et al. Reductions in dietary energy density are associated with weight loss in overweight and obese participants in the PREMIER trial. Am J Clin Nutr 2007;85:1212-21.
31. Kinross JM, von Roon AC, Holmes E, Darzi A, Nicholson JK. The human gut microbiome: implications for future health care. Curr Gastroenterol Rep 2008;10:396-403.
32. Chen L, Appel LJ, Loria C, et al. Reduction in consumption of sugar-sweetened beverages is associated with weight loss: the PREMIER trial. Am J Clin Nutr 2009;89:1299-306.
33. Schulze MB, Manson JE, Ludwig DS,

et al. Sugar-sweetened beverages, weight gain, and incidence of type 2 diabetes in young and middle-aged women. JAMA 2004;292:927-34.
34. Liu S, Serdula MK, Williamson DF, Mokdad AH, Byers T. A prospective study of alcohol intake and change in body weight among US adults. Am J Epidemiol 1994;140:912-20.
35. Wannamethee SG, Shaper AG. Alcohol, body weight, and weight gain in middle-aged men. Am J Clin Nutr 2003; 77:1312-7.
36. Wannamethee SG, Field AE, Colditz GA, Rimm EB. Alcohol intake and 8-year weight gain in women: a prospective study. Obes Res 2004;12:1386-96.
37. Wang L, Lee IM, Manson JE, Buring JE, Sesso HD. Alcohol consumption, weight gain, and risk of becoming overweight in middle-aged and older women. Arch Intern Med 2010;170:453-61.
38. Stull AJ, Apolzan JW, Thalacker-Mercer AE, Iglay HB, Campbell WW. Liquid and solid meal replacement products differentially affect postprandial appetite and food intake in older adults. J Am Diet Assoc 2008;108:1226-30.
39. Lloyd-Jones D, Adams RJ, Brown TM, et al. Heart disease and stroke statistics — 2010 update: a report from the American Heart Association. Circulation 2010; 121(7):e46-e215. [Erratum, Circulation 2010;121(12):e260.]
40. Rajpathak SN, Rimm EB, Rosner B, Willett WC, Hu FB. Calcium and dairy intakes in relation to long-term weight gain in US men. Am J Clin Nutr 2006;83:559-66.
41. Rosell M, Håkansson NN, Wolk A. Association between dairy food consumption and weight change over 9 y in 19,352 perimenopausal women. Am J Clin Nutr 2006;84:1481-8.
42. Tremblay A, Gilbert JA. Milk products, insulin resistance syndrome and type 2 diabetes. J Am Coll Nutr 2009;28:Suppl 1: 91S-102S.
43. Tong X, Dong JY, Wu ZW, Li W, Qin LQ. Dairy consumption and risk of type 2 diabetes mellitus: a meta-analysis of cohort studies. Eur J Clin Nutr 2011 May 11 (Epub ahead of print).
44. Lanou AJ, Barnard ND. Dairy and weight loss hypothesis: an evaluation of the clinical trials. Nutr Rev 2008;66:272-9.
45. Dove ER, Hodgson JM, Puddey IB, Beilin LJ, Lee YP, Mori TA. Skim milk compared with a fruit drink acutely reduces appetite and energy intake in overweight men and women. Am J Clin Nutr 2009;90:70-5.
46. Coon KA, Goldberg J, Rogers BL, Tucker KL. Relationships between use of television during meals and children's food consumption patterns. Pediatrics 2001;107(1):E7.
47. Vereecken CA, Todd J, Roberts C, Mulvihill C, Maes L. Television viewing behaviour and associations with food habits in different countries. Public Health Nutr 2006;9:244-50.
48. Utter J, Scragg R, Schaaf D. Associations between television viewing and consumption of commonly advertised foods among New Zealand children and young adolescents. Public Health Nutr 2006;9: 606-12.
49. Temple JL, Giacomelli AM, Kent KM, Roemmich JN, Epstein LH. Television watching increases motivated responding for food and energy intake in children. Am J Clin Nutr 2007;85:355-61.
50. Dubois L, Farmer A, Girard M, Peterson K. Social factors and television use during meals and snacks is associated with higher BMI among pre-school children. Public Health Nutr 2008;11:1267-79.
51. Manios Y, Kondaki K, Kourlaba G, Grammatikaki E, Birbilis M, Ioannou E. Television viewing and food habits in toddlers and preschoolers in Greece: the GENESIS study. Eur J Pediatr 2009;168: 801-8.
52. Escobar-Chaves SL, Markham CM, Addy RC, Greisinger A, Murray NG, Brehm B. The Fun Families Study: intervention to reduce children's TV viewing. Obesity (Silver Spring) 200100;18:Suppl 1:S99-S101.
53. 2010 Sleep in America poll. Washington, DC: National Sleep Foundation, 2010.
54. Office on Smoking and Health National Center for Chronic Disease Prevention and Health Promotion. Women and smoking: a report of the Surgeon General. Atlanta: Centers for Disease Control and Prevention, March 2001.
55. Willi C, Bodenmann P, Ghali WA, Faris PD, Cornuz J. Active smoking and the risk of type 2 diabetes: a systematic review and meta-analysis. JAMA 2007;298:2654-64.
56. U.S. Department of Agriculture, Agricultural Research Service. USDA national nutrient database for standard reference. Release 23. 2010. (http://www.ars.usda.gov/ba/bhnrc/ndl.)
57. Plachta-Danielzik S, Landsberg B, Bosy-Westphal A, Johannsen M, Lange D, Müller MJ. Energy gain and energy gap in normal-weight children: longitudinal data of the KOPS. Obesity (Silver Spring) 2008;16:777-83.
58. Zhai F, Wang H, Wang Z, Popkin BM, Chen C. Closing the energy gap to prevent weight gain in China. Obes Rev 2008;9: Suppl 1:107-12.

Close Reading #5: Research Review Article

Addressing Driver Aggression: Contributions From Psychological Science

Authors: Christine M. Wichens, Robert E. Mann, David L. Wiesenthal, University of Toronto

Think before you read: Has aggressive driving affected you or someone you know? What were the circumstances, causes, and consequences of the incident(s)? Could anything have been done to prevent them?

Close Reading Exercise: Skim this article and notice how it is organized and written, and how the author synthesizes and cites information. You may take notes in the margins. As you read each section, answer the following questions. You do not have to read every detail of the entire article to be able to answer the questions. Skim for relevant information. When you finish, discuss your answers in groups of 3-4.

1. Summarize the nature of the paradox referred to in paragraph 1 of the introduction ("Aggressive roadway behavior increases the risk…"). Cite the authors of the review article (Wickens, etc.) in your summary.

   _____
   _____
   _____

2. Paraphrase the definition of driver aggression given in paragraph 2:
   "For the purposes of streamlining the current review of a vast literature, driver aggression will be used to refer to violations of highway traffic laws (e.g., speeding, tailgating, reckless driving) and less serious anger expressions (e.g., swearing, obscene gestures) that are assumed to result from hostility toward another motorist…"

   _____
   _____
   _____

3. Summarize the main Person-Related Contributors to driver aggression (pp. 387-388). Include a citation of the article.

   _____
   _____
   _____

4. Summarize the Situational-Related Contributors to driver aggression (pp. 388-389). Include a citation of the article.

   _____
   _____
   _____

5. What are some methods to alleviate aggressive driver behavior? Summarize the strategies in your own words, citing the article.

   _____
   _____

Wickens, C. M., Mann, R. E., & Wiesenthal, D. L. (2013). Addressing driver aggression: Contributions from psychological science. Current Directions in Psychological Science, 22(5), 386-391.

Addressing Driver Aggression: Contributions from Psychological Science

Christine M. Wickens[1], Robert E. Mann[1], and David L. Wiesenthal[2]
[1] Centre for Addiction and Mental Health, Toronto, Canada
[2] York University, Toronto, Canada

Abstract

Aggressive roadway behavior contributes to motor-vehicle collisions, resulting in significant injuries, fatalities, and related financial costs. Psychological models have identified person- and situation-related variables that are predictive of driver aggression, and these have been used to develop strategies to alleviate aggressive roadway behavior. Future psychological research directions are discussed.

Introduction

Aggressive roadway behavior increases the risk of motor-vehicle collisions (MVCs) and is associated with greater injury severity resulting from such collisions (Galovski, Malta, & Blanchard, 2006; Paleti, Eluru, & Bhat, 2010). Although estimates of the prevalence of aggressive driving vary considerably (see Galovski et al., 2006), the AAA Foundation for Traffic Safety (2009) reported that 56% of fatal crashes in the United States from 2003 through 2007 involved at least one driver action that is typically associated with driver aggression, such as excessive speeding or reckless/careless driving. Although 78% of Americans recognize the danger and resulting health and financial impact of aggressive driving, a significant number of American drivers admit to speeding to beat a yellow light (58%), pressuring other motorists to speed up (26%), and tailgating (22%; AAA Foundation for Traffic Safety, 2009). Given this paradox of attitude versus behavior, psychological science clearly has a role to play in furthering our understanding of what factors contribute to aggressive driver behavior and identifying potential solutions to the problem.

Defining Aggressive Driver Behavior

In addressing the issue of driver aggression, the first step must be to define the term. Most available statistics, including those cited in the prior paragraph, are based on a broad interpretation of aggressive driving; however, there has been controversy concerning which aggressive acts meet inclusion criteria. Many researchers have argued that the aggressive action must be deliberate. If one motorist has an accidental lapse in judgment and does not leave enough space when pulling in front of another driver, is this an example of aggressive driving? Another definitional issue involves the nature of the intention. Must the driver be motivated by hostility toward another motorist to be considered aggressive, or can the driver be motivated by impatience or an attempt to save time? Some researchers have argued that there is a distinction between aggressive and risky driving. The former involves harmful intent directed toward another motorist, whereas the latter involves exclusively selfish motives such as time urgency or thrill-seeking (for a thorough review of this debate, see Galovski et al., 2006; Wiesenthal, Lustman, & Roseborough, in press). For the purposes of streamlining the current review of a vast literature, *driver aggression* will be used to refer to violations of highway traffic laws (e.g., speeding, tailgating, reckless driving) and less serious anger expressions (e.g., swearing, obscene gestures) that are assumed to result from hostility directed toward another motorist; *driver violence* will be used to refer to violations of criminal laws (e.g., threatening harm, assault).

These acts are not errors or lapses in judgment; they are aberrant driving behaviors (see Reason,

Manstead, Stradling, Baxter, & Campbell, 1990) motivated specifically by hostility. Psychologists have postulated many theoretical models explaining driver aggression that hypothesize a combination of person-related and situational variables (e.g., Shinar, 1998).

Person-Related Contributors

Demographics

Person-related variables are those factors that are specific to the driver; arguably, they constitute the largest and most diverse class of contributory factors. Demographic characteristics are the most basic of these variables. Driver aggression is more common among the young and the unmarried, which may be explained by more frequent risk-taking behavior by these demographic groups. Driver aggression has also been seen more commonly among the well- educated and higher socioeconomic status groups, perhaps because they have more social engagements and may be more rushed for time, or they may be less deterred by the risk of fines should they be observed by the authorities (Wickens et al, 2012). Driver aggression has been shown to be greater among men than women, but the most significant gender difference is found with driver violence: men are much more likely to engage in this extreme behavior (Hennessy, Wiesenthal, Wickens, & Lustman, 2004).

Personality

Personality may affect our cognitive perception of a situation, our preferences regarding levels of arousal or stimulation, or our sensitivity to stress or threat, all of which play a role when we are driving (Matthews, Dorn, & Glendon, 1991). Drivers who frequently demonstrate high levels of verbal and physical aggression or anger in other aspects of their lives are generally more likely to do so in the driving environment (Deffenbacher, Deffenbacher, Lynch, & Richards, 2003). Narcissistic people are recognized as arrogant, selfish, and having a sense of entitlement. Narcissists have been found to engage in more retaliatory and vengeful behavior, perhaps because they are more likely to perceive ambiguous driving altercations as intentional or unjust (Lustman, Wiesenthal, & Flett, 2010). Sensation seeking, associated with a need for novel and intense stimuli, has generally been associated with risky driving behavior. This trait has also been identified as a significant predictor of driver aggression, perhaps because sensation seekers perceive less risk in, or accept the risk associated with, roadway aggression (Dahlen, Martin, Ragan, & Kuhlman, 2005). Impulsive individuals demonstrate poor control over thoughts and behaviors, often initiating behavior without significant forethought, and are more likely to use the vehicle as a weapon for retaliation (Dahlen et al., 2005). Type-A personality consists of a cluster of traits relevant to driver behavior including competitiveness, hostility, achievement motivation, and a sense of time urgency (Bone & Mowen, 2006; Wickens & Wiesenthal, 2005). Not surprisingly, Type-A personality is more common among aggressive than nonaggressive drivers (Miles & Johnson, 2003). Neuroticism is associated with feelings of anxiety, anger, envy, depressed mood, and poor emotional response to stress. Drivers high in neuroticism engage in more horn honking, tailgating, and using obscene hand gestures (Bone & Mowen, 2006). Other variables that have been found to contribute to driver aggression include machismo, extraversion, ego defensiveness, and emotional instability (Bone & Mowen, 2006; Krahé & Fenske, 2002; Neighbors, Vietor, & Knee, 2002; Sümer, Lajunen, & Özkan, 2005). There are also personality variables that have been found to reduce the likelihood that a driver will engage in roadway aggression, including high levels of conscientiousness and agreeableness (Bone & Mown, 2006; Sümer et al., 2005).

Cognition

How we cognitively perceive a driving event will have a major impact on how we feel and eventually respond to the event. Stress researchers conceptualize cognition in driver aggression as involving appraisal of the demands of a stressful situation and ability to cope with them. A driver caught in a stressful driving situation characterized by crowded but quickly

moving traffic, time urgency, and an unexpected near-collision may assess the situation as being greater than his/her personal resources can tolerate. The motorist may experience feelings of anger and may lash out aggressively (Matthews et al., 1991; Wickens & Wiesenthal, 2005). Attribution theorists have conceptualized the role of cognition as a series of judgments regarding why an event occurred and the level of responsibility assigned to an offending driver. If we are cut off on the highway and assume that the offending driver's actions were intentional, we feel angry and may respond in kind. However, if we attribute the driver's actions to an unintentional cause such as a sudden tire blowout causing the vehicle to swerve in front of us, then we may feel sympathy for the other motorist (Wickens, Wiesenthal, Flora, & Flett, 2011).

Cognitive biases can also influence the development of driver aggression. When interpreting the potentially offensive actions of other motorists, drivers tend to overestimate internal (e.g., personality) and underestimate external (e.g., situation) causes; however, drivers tend to do the opposite when making attributions for their own actions (i.e., the actor-observer bias; e.g., Herzog, 1994). Novice motorists tend to be overconfident of their driving skills (Mynttinen et al., 2009), thus lowering their tolerance for the perceived misdeeds of other motorists.

Alcohol, drugs, and mental health

Alcohol-related problems, use of cannabis, and use of these substances immediately before driving increase one's risk of engaging in driver aggression (Butters, Mann, & Smart, 2006; Wickens et al., 2012). Drivers reporting the use of cocaine, ecstasy (MDMA), or both are more likely to commit violent roadway behavior (Butters et al., 2006). The pharmacological effect of these substances on mood and inhibition, along with personality characteristics (e.g., trait anger or aggression, sensation seeking) common to drinkers, drug users, and aggressive drivers, may also explain the overlap in these behaviors.

Various psychiatric disorders have also been implicated as contributors to driver aggression. Intermittent explosive disorder is an impulse control disorder characterized by extreme expressions of anger out of proportion to the provoking stimulus. In a study of treatment-seeking aggressive drivers in Albany, New York, approximately one third of these drivers met criteria for intermittent explosive disorder, significantly more than a control sample of nonaggressive drivers (Galovski et al., 2006). Attention deficit hyperactivity disorder is characterized by inattention, impulsivity, and hyperactivity and is associated with increased self-reports of driving violations, anger, and aggression (Barkley & Cox, 2007). Attention deficit hyperactivity disorder often co-occurs with other disruptive behavior disorders, such as conduct disorder and oppositional defiant disorder. Relative to a sample of nonaggressive control subjects, these disorders have been found to be more prevalent among aggressive drivers (Malta, Blanchard, & Freidenberg, 2005). Personality disorders, such as antisocial personality disorder and paranoid personality disorder, are also more likely to be found among aggressive than nonaggressive drivers (Galovski et al., 2006). Psychiatric distress, which includes symptoms of both depression and anxiety, has been found to significantly increase the odds of perpetrated driver violence (Butters et al., 2006). Nonetheless, studies examining the impact of anxiety and mood disorders on driver aggression have generated mixed findings, providing some support for this relationship but necessitating additional research (Wickens, Mann, Butters, Smart, & Stoduto, in press). Finally, it is also important to note that medications used to ameliorate psychiatric problems may influence, and perhaps increase, driver aggression (Wickens, Mann, Butters, et al., in press).

Situation-Related Contributors

Environmental factors

Sights, sounds, and smells can all play a role. The visual content of the roadside environment can influence the level of stress and negative affect experienced by drivers; urban

roadways lined with commercial buildings and billboards generate more stress than rural roadways lined with natural vegetation (Parsons, Tassinary, Ulrich, Hebl, & Grossman-Alexander, 1998). Likewise, hostile cues such as aggressive billboard advertising or a gun rack in the window of a pickup truck increase driver anger and aggression (Ellison-Potter, Bell, & Deffenbacher, 2001). Sounds within the vehicle can also influence stress levels; self-selected music reduces stress experienced in heavy traffic congestion (Wiesenthal, Hennessy, & Totten, 2000). Likewise, the smell of peppermint decreases drivers' frustration, anxiety, and fatigue (Raudenbush, Grayhem, Sears, & Wilson, 2009), and rising ambient temperature increases drivers' horn-honking (Kenrick & MacFarlane, 1986).

Situational factors

Within the driving environment, aspects of the situation can also elicit or augment anger behind the wheel that would not otherwise have emerged. Offensive driving by another motorist can provoke roadway anger and aggression (Wickens et al., 2011), but situational factors can further increase the likelihood of an aggressive response. Traffic congestion is a major source of roadway stress and the resulting frustration may be directed aggressively at other motorists (Shinar, 1998). Daily hassles and job-related stresses can make traffic congestion or an offensive driver action seem much more upsetting (Matthews et al., 1991; Wickens & Wiesenthal, 2005). Likewise, time urgency can make traffic congestion or an otherwise benign traffic situation seem much more stressful (Wickens & Wiesenthal, 2005), which can lead to driver aggression.

Attributions of other drivers' roadway actions are influenced by the visible characteristics of that driver and the features of their vehicle. Female drivers are judged to be more careless and less aggressive than male drivers, and drivers of BMWs are judged to be more aggressive than drivers of Smart Cars (Lawrence & Richardson, 2005). The relative status of vehicles also makes a difference in the likelihood of aggression; when blocked by a "middle-class" vehicle stopped at a green light, drivers of upper-class vehicles honk their horns more quickly than drivers of middle-class vehicles, who honk more quickly than drivers of lower-class vehicles (Diekmann, Jungbauer-Gans, Krassnig, & Lorenz, 1996).

Alleviating Aggressive Driver Behavior

Beyond bettering our understanding of the factors that contribute to driver violence and aggression, psychological science is also developing strategies to alleviate the behavior. Programs to treat aggressive drivers are now being developed using cognitive-behavioral therapy, attributional retraining, and relaxation training (Galovski et al., 2006). These programs teach drivers to identify the triggers of their roadway anger and aggression, to recognize cognitive distortions that contribute to their anger, and to control their breathing and relax their muscles when an anger-provoking event is encountered. Additional evidence-based curricula could be added, such as recognizing the tendency to overestimate our own driving skills and emphasizing the importance of roadway communication (e.g., signaling lane changes, flashing headlights as a sign of gratitude; Wickens et al., 2011). Although development of these programs is in the early stages, the success of similarly-intended programs for persons convicted of driving while intoxicated (e.g., Wickens, Mann, Stoduto, Flam Zalcman, & Butters, 2013) suggests that these programs could substantially improve traffic safety if implemented on a large scale. These programs could also be beneficial if presented early in a novice driver's training.

Other attempts at behavior modification have included incentives for good driving: Instrumented vehicles or monitored traffic zones identify and reward law-abiding drivers with entries in a lottery or direct monetary compensation (Battista, Burns, & Taylor, 2010; Haggarty, 2010). Directed passenger feedback has also been used to encourage drivers to better monitor their speed and mirrors (Hutton, Sibley, Harper, & Hunt, 2002), although it is unclear how long this effect might persist, whether it could be used to reduce retaliatory aggression, and whether it is affected by the type of relationship between the driver and the passenger (e.g., teen driver and parent; see Wiesenthal et al., in press).

Psychological science can also advise police by identifying specific driving behaviors for enforcement campaigns and when these efforts should occur (Wickens, Wiesenthal, Hall, & Roseborough, 2013). It can inform public service and education campaigns through identification of the audience to target and the most effective focus of the public appeal (e.g., emotional versus informational; Lewis, Watson, White, & Tay, 2007). Psychological science also allows for the evaluation of various technological solutions to the driver aggression problem including photo radar, red-light cameras, and electronic message boards over the highway for safety appeals or in the rear window of a passenger vehicle to facilitate inter-vehicle communication (e.g., Chen, Meckle, & Wilson, 2002; Retting, Williams, Farmer, & Feldman, 1999; Smart, Cannon, Howard, & Mann, 2005).

Future Directions

Psychology is advancing our knowledge of factors contributing to driver aggression, adding to the list of relevant variables and expanding our understanding of existing factors. Person-related and situational variables operate together; thus, it is imperative that we continue to investigate how the contributions of multiple factors combine and interact to influence aggressive roadway behavior. We also need to understand the mechanisms underlying the influence of contributory factors. Personality, cognition, and affect all influence each other, and an improved assessment of the temporal order and strength of these influences is needed. Efforts to apply this information to modify driver aggression through policy, incentive-based approaches, psychotherapeutic programs (e.g., attributional retraining), and technological innovations to the vehicle and the roadway environment (e.g., electronic message boards) are in their infancy but possess great potential for impact.

Acknowledgments

The authors express their gratitude for the valuable comments provided by three anonymous reviewers.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Funding

This work was supported by a grant from AUTO21, a member of the Networks of Centres of Excellence program that is administered and funded by the Natural Sciences and Engineering Research Council, the Canadian Institutes of Health Research, and the Social Sciences and Humanities Research Council (SSHRC), in partnership with Industry Canada. C. M. Wickens was supported by Postdoctoral Fellowships from the SSHRC and the Centre for Addiction and Mental Health, and R. E. Mann acknowledges ongoing funding support from the Ontario Ministry of Health and Long-Term Care.

References

AAA Foundation for Traffic Safety. (2009, April). *Aggressive driving: Research update.* Washington, DC: AAA Foundation for Traffic Safety.

Barkley, R. A., & Cox, D. (2007). A review of driving risks and impairments associated with attention-deficit/hyperactivity disorder and the effects of stimulant medication on driving performance. *Journal of Safety Research, 38,* 113-28.

Battista, V., Burns, P., & Taylor, G. (2010). Using rewards to influence driving behaviour: A field operational trial. *Proceedings of the 20th Canadian Multidisciplinary Road Safety Conference, Niagara Fall, Ontario, Canada*.

Bone, S. A., & Mowen, J. C. (2006). Identifying the traits of aggressive and distracted drivers: a

hierarchical trait model approach. *Journal of Consumer Behaviour, 5,* 454-464.
Butters, J. E., Mann, R. E., & Smart, R. G. (2006). Assessing road rage victimization and perpetration in the Ontario adult population. *Canadian Journal of Public Health, 97,* 96-99.
Chen, G., Meckle, W., & Wilson, J. (2002). Speed and safety effect of photo radar enforcement on a highway corridor in British Columbia. *Accident Analysis and Prevention, 34,* 129-138.
Dahlen, E. R., Martin, R. C., Ragan, K., & Kuhlman, M. M. (2005). Driving anger, sensation seeking, impulsiveness, and boredom proneness in the prediction of unsafe driving. *Accident Analysis and Prevention, 37,* 341-348.
Deffenbacher, J. L., Deffenbacher, D. M., Lynch, R. S., & Richards, T. L. (2003). Anger, aggression, and risky behavior: a comparison of high and low anger drivers. *Behaviour Research and Therapy, 41,* 701-718.
Diekmann, A., Jungbauer-Gans, M., Krassnig, H., & Lorenz, S. (1996). Social status and aggression: A field study analyzed by survival analysis. *The Journal of Social Psychology, 136,* 761-768.
Ellison-Potter, P., Bell, P., & Deffenbacher, J. (2001). The effects of trait driving anger, anonymity, and aggressive stimuli on aggressive driving behavior. *Journal of Applied Social Psychology, 31,* 431-443.
Galovski, T. E., Malta, L. S., Blanchard, E. B. (2006). *Road rage: Assessment and treatment of the angry, aggressive driver.* Washington, DC: American Psychological Association.
Haggarty, E. (2010, December 9). Speed camera lottery pays drivers for slowing down. *The Toronto Star*, December 9, 2010. Retrieved from www.thestar.com/printarticle/904004.
Hennessy, D. A., Wiesenthal, D. L., Wickens, C. M., & Lustman, M. (2004). The impact of gender and stress on traffic aggression: Are we really that different?. In J. P. Morgan (Ed.), *Focus on aggression research* (pp. 157-174). Hauppauge, NY: Nova Science Publishers.
Herzog, T. A. (1994). Automobile driving as seen by the actor, the active observer, and the passive observer. *Journal of Applied Social Psychology, 24,* 2057-2074.
Hutton, K. A., Sibley, C. G., Harper, D. N., & Hunt, M. (2002). Modifying driver behavior with passenger feedback. *Transportation Research Part F*, *4*, 257–269.
Kenrick, D. T., & MacFarlane, S. W. (1986). Ambient temperature and horn honking: A field study of the heat/aggression relationship. *Environment and Behavior, 18,* 179-191.
Krahé, B., & Fenske, I. (2002). Predicting aggressive driving behavior: The role of macho personality, age, and power of car. *Aggressive Behavior, 28,* 21-29.
Lawrence, C., & Richardson, J. (2005). Gender-based judgments of traffic violations: The moderating influence of car type. *Journal of Applied Social Psychology, 35,* 1755-1774
Lewis, I. M., Watson, B., White, K. M., & Tay, R. (2007). Promoting public health messages: Should we move beyond fear-evoking appeals in road safety? *Qualitative Health Research, 17,* 61-74.
Lustman, M., Wiesenthal, D. L., & Flett, G. L. (2010). Narcissism and aggressive driving: Is an inflated view of the self a road hazard? *Journal of Applied Social Psychology, 40,* 1423-1449.
Malta, L. S., Blanchard, E. B., & Freidenberg, B. M. (2005). Psychiatric and behavioral problems in aggressive drivers. *Behaviour Research and Therapy, 43,* 1467-84.
Matthews, G., Dorn, L., & Glendon, A. I. (1991). Personality correlates of driver stress. *Personality and Individual Differences, 12,* 535-549.
Miles, D. E., & Johnson, G. L. (2003). Aggressive driving behaviors: are there psychological and attitudinal predictors? *Transportation Research Part F, 6,* 147-161.
Mynttinen, S., Sundström, A., Koivukoski, M., Hakuli, K., Keskinen, E., & Henriksson, W. (2009). Are novice drivers overconfident? A comparison of self-assessed and examiner-assessed driver competences in a Finnish and Swedish sample. *Transportation Research Part F, 12,* 120-130.

Neighbors, C., Vietor, N. A., & Knee, C. R. (2002). A motivational model of driving anger and aggression. *Personality and Social Psychology Bulletin, 28,* 324-335.

Paleti, R., Eluru, E., & Bhat, C. R. (2010). Examining the influence of aggressive driving behavior on driver injury severity in traffic crashes. *Accident Analysis and Prevention, 42,* 1839-1854.

References

Bates College. (2011). “The Structure, Format, Content, and Style of a Journal-Style Scientific Paper. http://www.abacus.bates.edu/~ganderso/biology/resources/writing/HTWsections.html

Brendgen, Mara & Dionne, Ginette & Girard, Alain & Boivin, Michel & Vitaro, Frank & Pérusse, Daniel. (2005). Examining Genetic and Environmental Effects on Social Aggression: A Study of 6-Year-Old Twins. Child development. 76. 930-46.

Gilmanov, Timur. (2018). Lower bound resource requirements for machine intelligence. Doctoral Dissertation. Indiana University.

Graff, Gerald and Cathy. Birkenstein. (2012). "They Say / I Say": The Moves That Matter in Academic Writing: with Readings. New York: W.W. Norton & Co.

Haenssle, Holger & Fink, Christine & Schneiderbauer, R & Toberer, Ferdinand & Buhl, Timo & Blum, A & Kalloo, A & Ben Hadj Hassen, A & Thomas, Luc & Enk, A & Uhlmann, Lorenz. (2018). Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists. Annals of oncology: official journal of the European Society for Medical Oncology. 29.

Henderson, Eric. (2007). The Active Reader: Strategies for Academic Reading and Writing. Toronto: Oxford University Press. Print.

Lightman, Bernard. (2010). “The Microscopic World.” Victorian Review. 36.

Mozaffarian, D., Hao, T., Rimm, E. B., Willett, W. C., & Hu, F. B. (2011). Changes in diet and lifestyle and longterm weight gain in women and men. *The New England journal of medicine*, *364*(25), 2392–2404.

Reseapro Scientific Services. (2015). “Journal Article vs Thesis.” *Publication Manual.* https://www.pubmanu.com/journal-article-vs-thesis/

Wickens, C. M., Mann, R. E., & Wiesenthal, D. L. (2013). Addressing driver aggression: Contributions from psychological science. *Current Directions in Psychological Science, 22*(5), 386-391.