Nonparametric Statistics

Kin 304W

Week 8: June 26, 2012

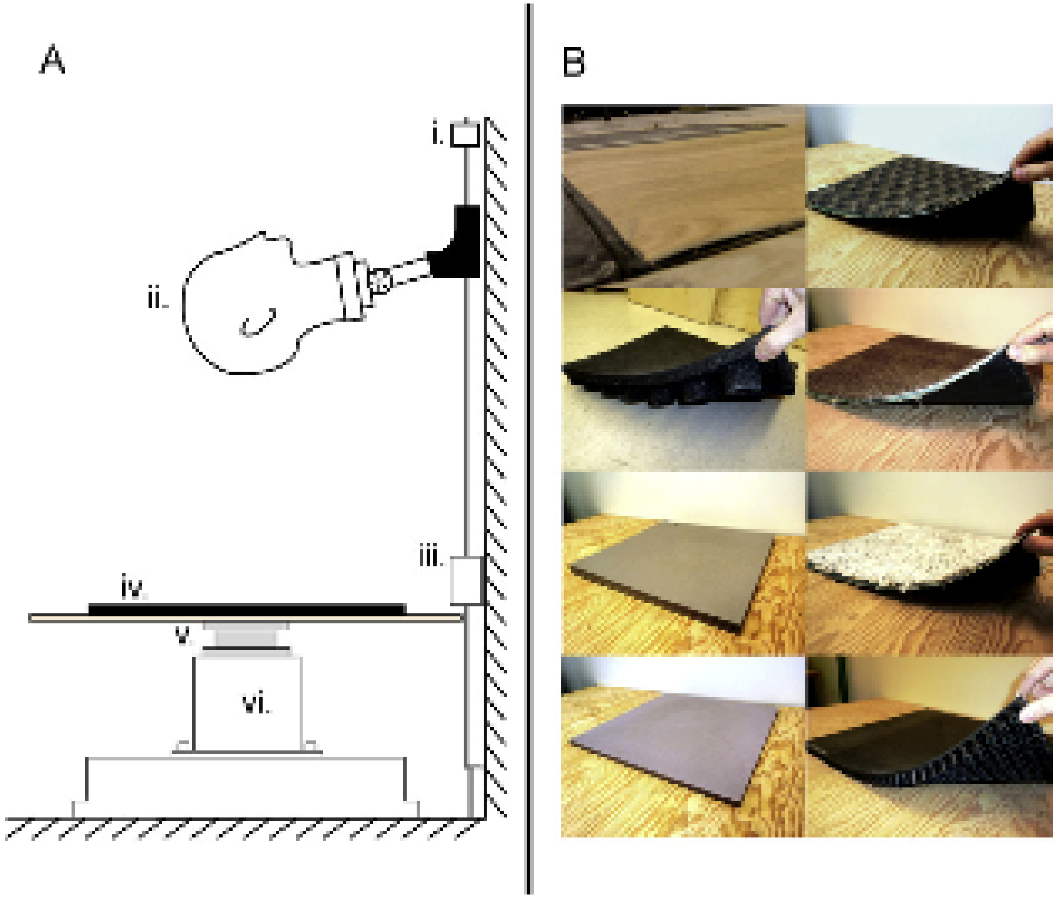
Today's Outline

- Project
  - Organization of Introduction, Methods, & Discussion sections
  - We'll take up an example article:

    Wright AD, Laing AC. The influence of headform orientation and flooring systems on impact dynamics during simulated fall-related head impacts. Med Eng Phys (2011), doi:10.1016/j.medengphy.2011.11.012

- Writer's Corner: Active vs. Passive Voice
- Nonparametric statistics

Example Article: Wright AD, Laing AC. 2011. The influence of headform orientation and flooring systems on impact dynamics during simulated fall-related head impacts.



Dependent variables:

Peak impact force (Fmax)
Peak linear acceleration (gmax)
Head Injury Criterion (HIC)

Independent Variables: Exp. #1:

Orientation (front, side, back)
Velocity (1.5, 2.5, 3.5 m/s)

Independent Variables: Exp. #2:

Floor condition (10 floors)
Velocity (1.5, 2.5, 3.5 m/s)

For each dependent variable, what type of analysis would you do for Exp #1? Exp #2?

Project: Introduction

- Start broadly, then narrow to a focal point - the research questions
- What is known about the topic?
  - Must reference previous research
- What is unknown (and needs to be known)?
  - This can often be written effectively in one sentence
- What is the research question?
  - "The purpose of this study was to test the hypothesis that…"
  - "This study sought to answer the following research questions:…"
- Why is this research topic/question important?
  - Why do we need to know the answer to the research questions?
  - Include statements about the importance of the health problem

Project: Introduction, Wright & Laing 2011

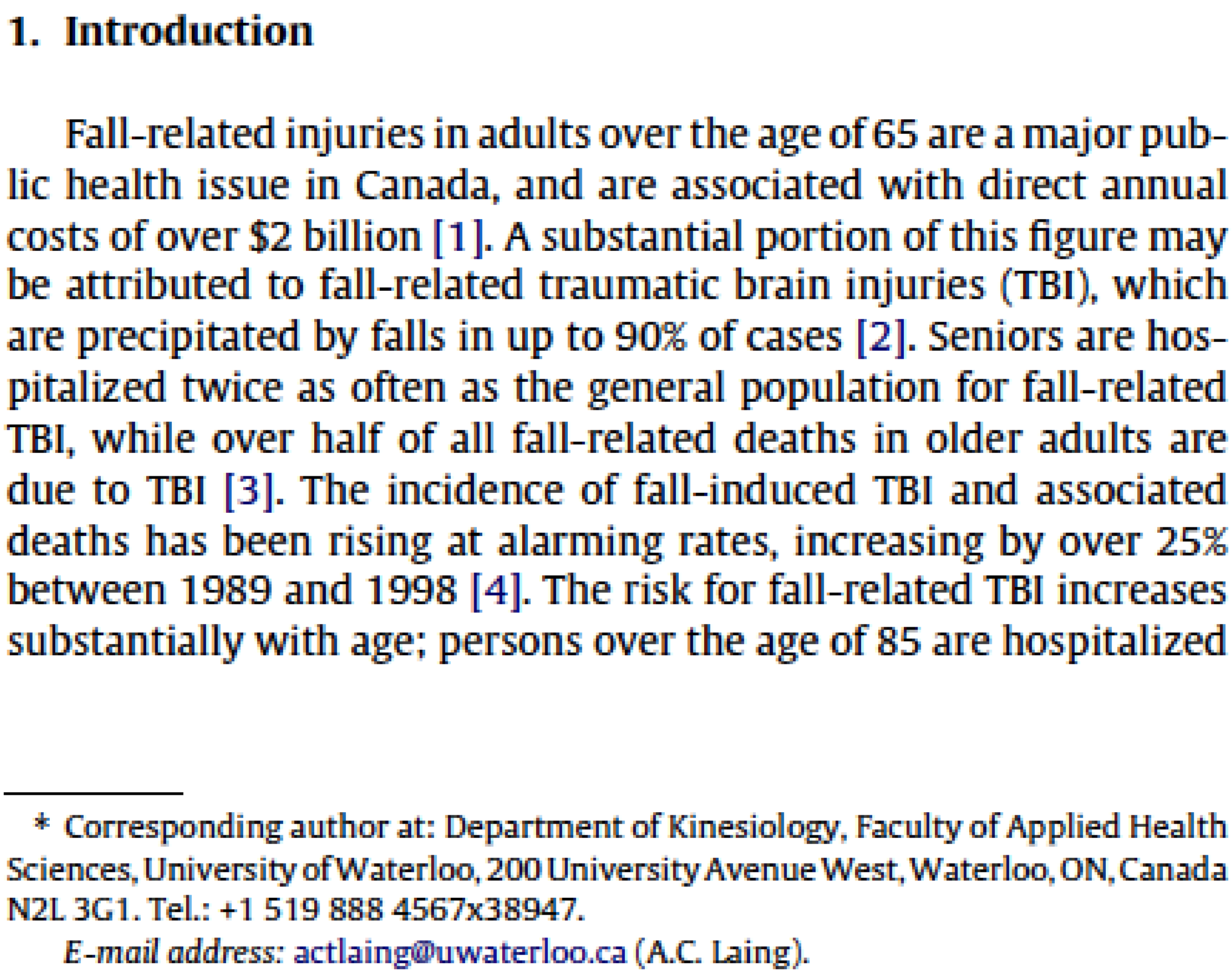
- Why is the topic important?

- What is known?

- What is unknown?

- What is the question?

Para 1



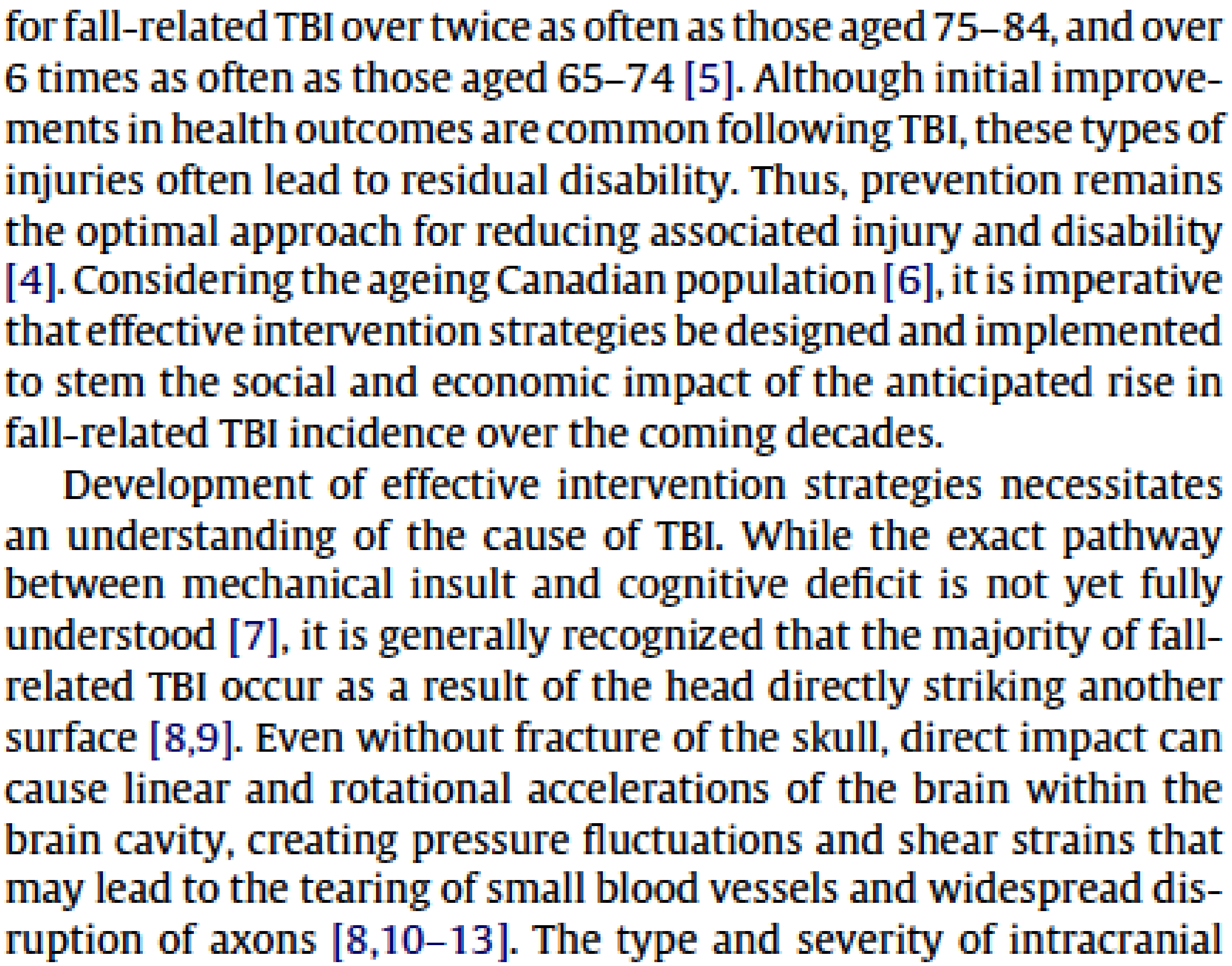
1. Introduction

Fall-related injuries in adults over the age of 65 are a major public health issue in Canada, and are associated with direct annual costs of over $2 billion [1]. A substantial portion of this figure may be attributed to fall-related traumatic brain injuries (TBI), which are precipitated by falls in up to 90% of cases [2]. Seniors are hospitalized twice as often as the general population for fall-related TBI, while over half of all fall-related deaths in older adults are due to TBI [3]. The incidence of fall-induced TBI and associated deaths has been rising at alarming rates, increasing by over 25% between 1989 and 1998 [4]. The risk for fall-related TBI increases substantially with age; persons over the age of 85 are hospitalized

* Corresponding author at: Department of Kinesiology, Faculty of Applied Health Sciences, University of Waterloo, 200 University Avenue West, Waterloo, ON, Canada N2L 3G1. Tel.: +1 519 888 4567x38947.

E-mail address: actlaing@uwaterloo.ca (A.C. Laing).

Please cite this article in press as: Wright AD, Laing AC. The influence of headform orientation and flooring systems on impact dynamics during simulated fall-related head impacts. Med Eng Phys (2011), doi:10.1016/j.medengphy.2011.11.012



Para 1 cont.

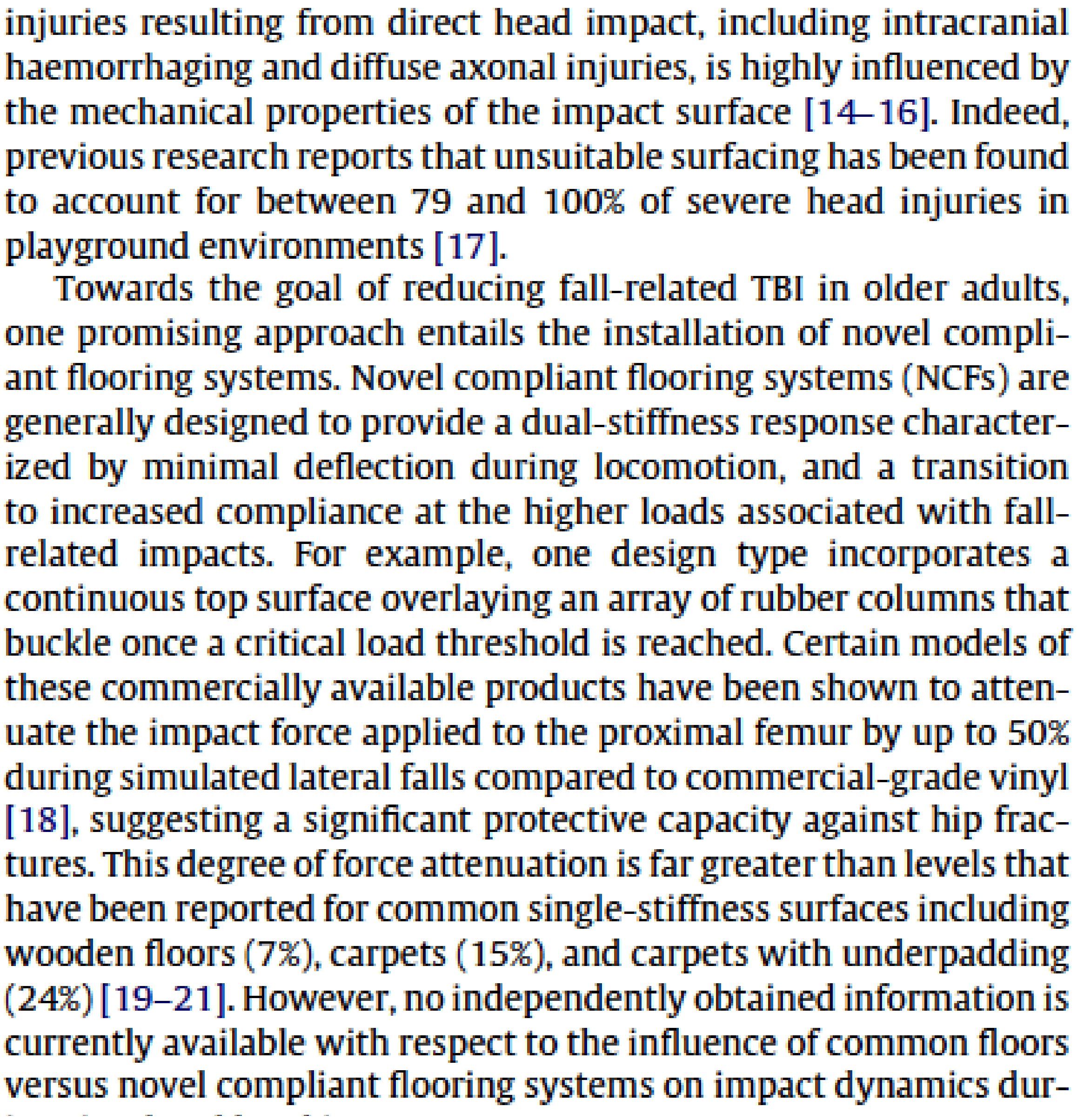
for fall-related TBI over twice as often as those aged 75–84, and over 6 times as often as those aged 65–74 [5]. Although initial improvements in health outcomes are common following TBI, these types of injuries often lead to residual disability. Thus, prevention remains the optimal approach for reducing associated injury and disability [4]. Considering the ageing Canadian population [6], it is imperative that effective intervention strategies be designed and implemented to stem the social and economic impact of the anticipated rise in fall-related TBI incidence over the coming decades.

Para 2

Development of effective intervention strategies necessitates an understanding of the cause of TBI. While the exact pathway between mechanical insult and cognitive deficit is not yet fully understood [7], it is generally recognized that the majority of fall-related TBI occur as a result of the head directly striking another surface [8,9]. Even without fracture of the skull, direct impact can cause linear and rotational accelerations of the brain within the brain cavity, creating pressure fluctuations and shear strains that may lead to the tearing of small blood vessels and widespread disruption of axons [8,10–13]. The type and severity of intracranial

Please cite this article in press as: Wright AD, Laing AC. The influence of headform orientation and flooring systems on impact dynamics during simulated fall-related head impacts. Med Eng Phys (2011), doi:10.1016/j.medengphy.2011.11.012

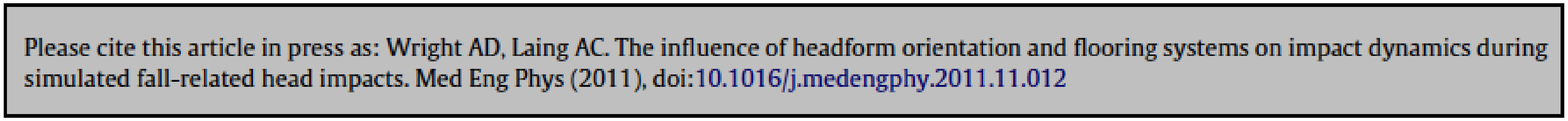
Para 2 cont.



injuries resulting from direct head impact, including intracranial haemorrhaging and diffuse axonal injuries, is highly influenced by the mechanical properties of the impact surface [14–16]. Indeed, previous research reports that unsuitable surfacing has been found to account for between 79 and 100% of severe head injuries in playground environments [17].

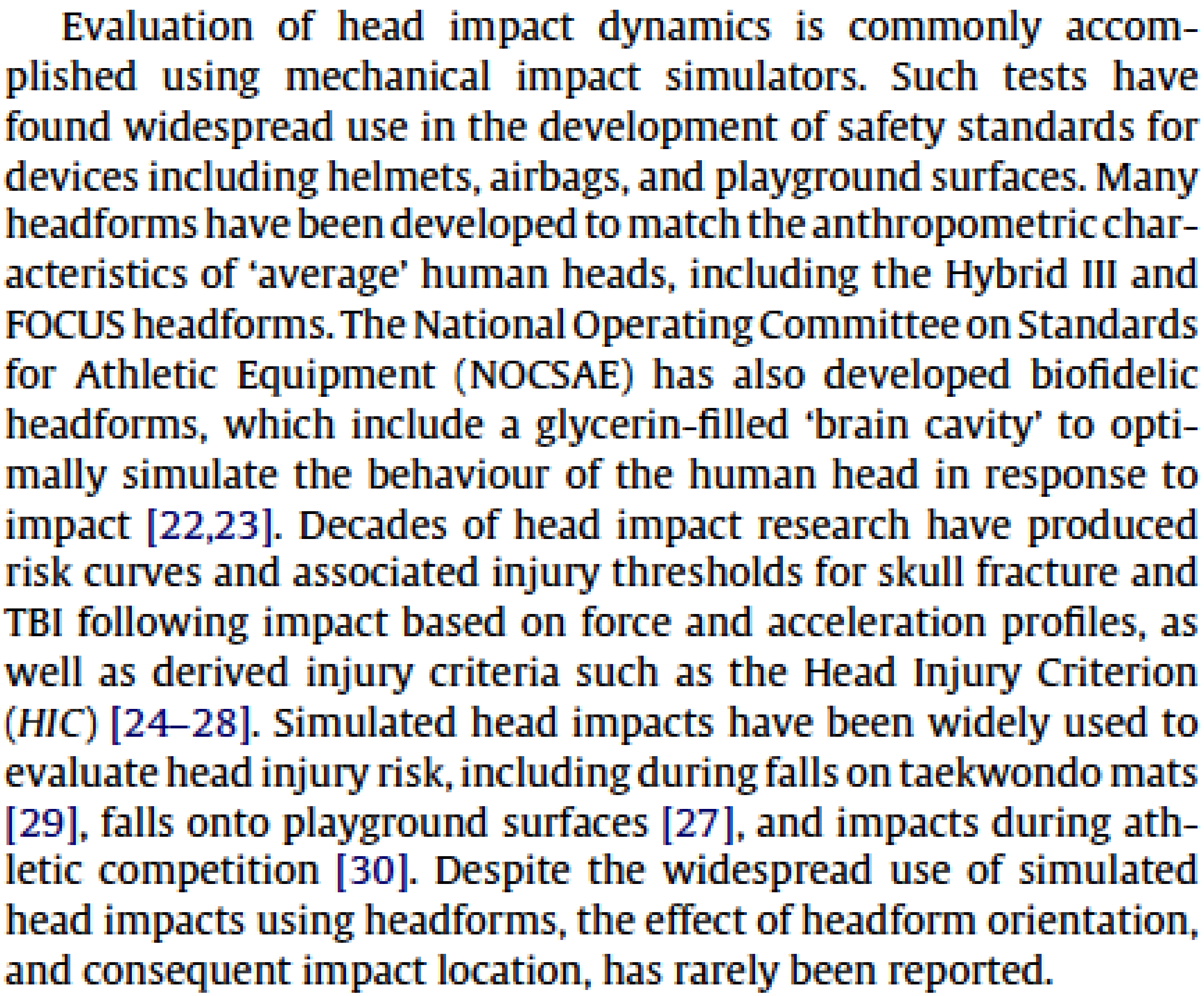
Para 3

Towards the goal of reducing fall-related TBI in older adults, one promising approach entails the installation of novel compliant flooring systems. Novel compliant flooring systems (NCFs) are generally designed to provide a dual-stiffness response characterized by minimal deflection during locomotion, and a transition to increased compliance at the higher loads associated with fall-related impacts. For example, one design type incorporates a continuous top surface overlaying an array of rubber columns that buckle once a critical load threshold is reached. Certain models of these commercially available products have been shown to attenuate the impact force applied to the proximal femur by up to 50% during simulated lateral falls compared to commercial-grade vinyl [18], suggesting a significant protective capacity against hip fractures. This degree of force attenuation is far greater than levels that have been reported for common single-stiffness surfaces including wooden floors (7%), carpets (15%), and carpets with underpadding (24%) [19–21]. However, no independently obtained information is currently available with respect to the influence of common floors versus novel compliant flooring systems on impact dynamics dur-

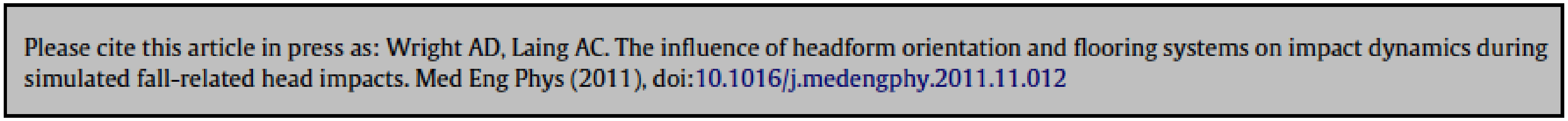


Please cite this article in press as: Wright AD, Laing AC. The influence of headform orientation and flooring systems on impact dynamics during simulated fall-related head impacts. Med Eng Phys (2011), doi:10.1016/j.medengphy.2011.11.012

Para 4

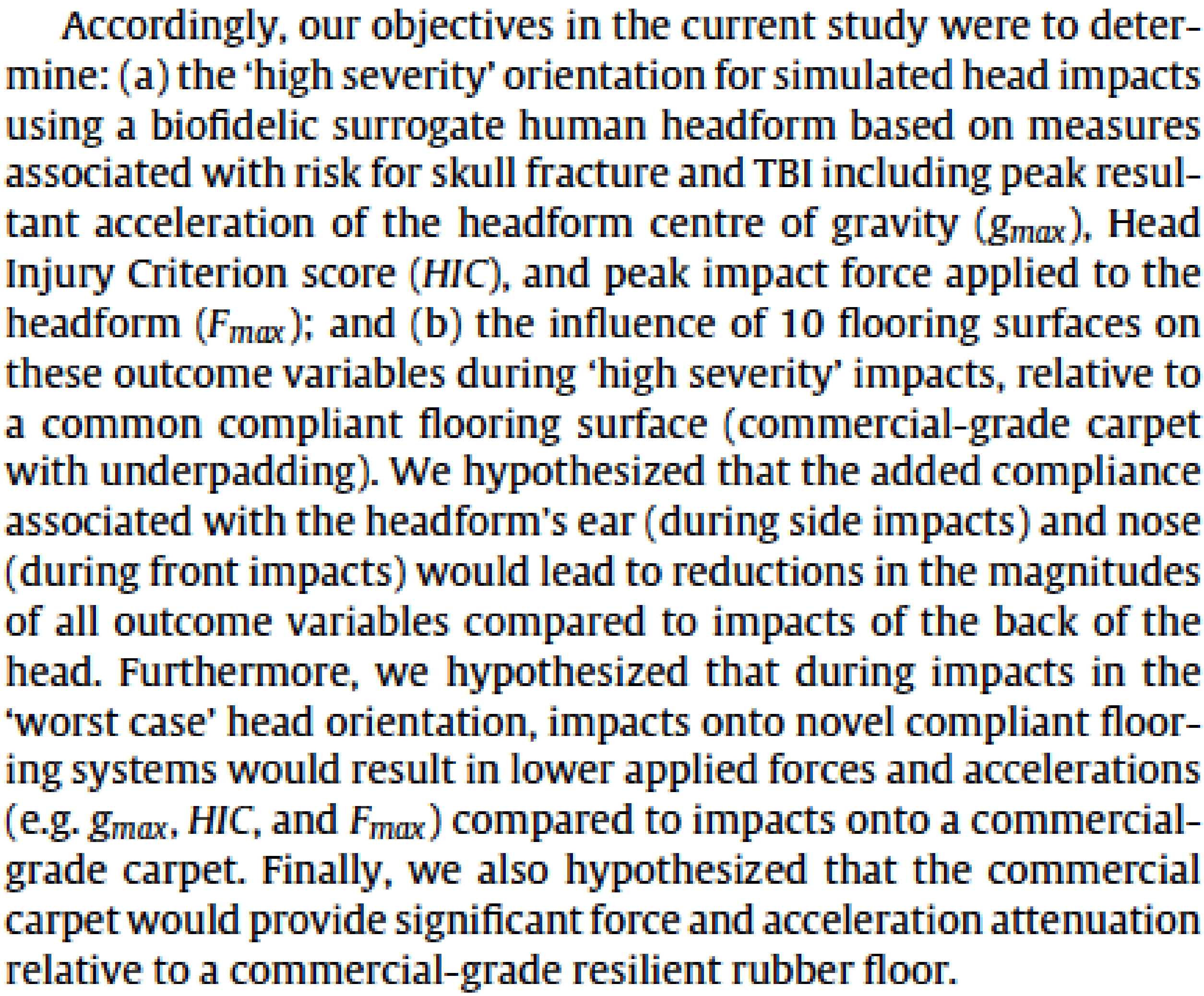


Evaluation of head impact dynamics is commonly accomplished using mechanical impact simulators. Such tests have found widespread use in the development of safety standards for devices including helmets, airbags, and playground surfaces. Many headforms have been developed to match the anthropometric characteristics of 'average' human heads, including the Hybrid III and FOCUS headforms. The National Operating Committee on Standards for Athletic Equipment (NOCSAE) has also developed biofidelic headforms, which include a glycerin-filled 'brain cavity' to optimally simulate the behaviour of the human head in response to impact [22,23]. Decades of head impact research have produced risk curves and associated injury thresholds for skull fracture and TBI following impact based on force and acceleration profiles, as well as derived injury criteria such as the Head Injury Criterion (*HIC*) [24–28]. Simulated head impacts have been widely used to evaluate head injury risk, including during falls on taekwondo mats [29], falls onto playground surfaces [27], and impacts during athletic competition [30]. Despite the widespread use of simulated head impacts using headforms, the effect of headform orientation, and consequent impact location, has rarely been reported.

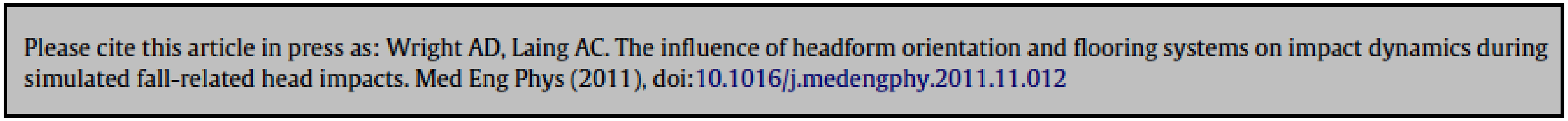


Please cite this article in press as: Wright AD, Laing AC. The influence of headform orientation and flooring systems on impact dynamics during simulated fall-related head impacts. Med Eng Phys (2011), doi:10.1016/j.medengphy.2011.11.012

Para 5



Accordingly, our objectives in the current study were to determine: (a) the 'high severity' orientation for simulated head impacts using a biofidelic surrogate human headform based on measures associated with risk for skull fracture and TBI including peak resultant acceleration of the headform centre of gravity (g_{max}), Head Injury Criterion score (*HIC*), and peak impact force applied to the headform (F_{max}); and (b) the influence of 10 flooring surfaces on these outcome variables during 'high severity' impacts, relative to a common compliant flooring surface (commercial-grade carpet with underpadding). We hypothesized that the added compliance associated with the headform's ear (during side impacts) and nose (during front impacts) would lead to reductions in the magnitudes of all outcome variables compared to impacts of the back of the head. Furthermore, we hypothesized that during impacts in the 'worst case' head orientation, impacts onto novel compliant flooring systems would result in lower applied forces and accelerations (e.g. g_{max}, *HIC*, and F_{max}) compared to impacts onto a commercial-grade carpet. Finally, we also hypothesized that the commercial carpet would provide significant force and acceleration attenuation relative to a commercial-grade resilient rubber floor.



Please cite this article in press as: Wright AD, Laing AC. The influence of headform orientation and flooring systems on impact dynamics during simulated fall-related head impacts. Med Eng Phys (2011), doi:10.1016/j.medengphy.2011.11.012

Project: Statistical Analysis

- Which statistical tests did you use and why?
  - e.g. We used independent *t* tests to compare waist to hip ratio and sum of five skinfolds between athletes and non-athletes.
- What statistical software package and version did you use?

Project: Statistical Analysis, Wright & Laing 2011

2.5. Statistics

2.5.1. Determination of the ‘high severity’ headform orientation

A two-way ANOVA was used to assess the influence of impact orientation and impact velocity on gmax , HIC, and Fmax . When significant interactions were found, simple effects were analyzed to determine the influence of impact orientation at each impact velocity, with Tukey’s post hoc used to compare across the three orientations.

2.5.2. Floor testing

A two-way ANOVA was used to assess the influence of floor condition and impact velocity on each of the outcome parameters. If a significant interaction was found, simple effects were analyzed to determine the influence of floor condition at each impact velocity. Dunnett’s post hoc test (which is appropriate when a baseline comparator condition exists) was used to compare each floor relative to the control condition, Carpet$_{\text{comm}}$.

To account for the use of three dependent variables, we used an alpha of 0.0167 (i.e. 0.05/3) for ANOVAs. Post hoc tests were conducted with an experiment-wide significance level of 0.05 using SPSS statistical software package (Version 19.0, SPSS Inc., Chicago, IL, USA).

Project: Discussion

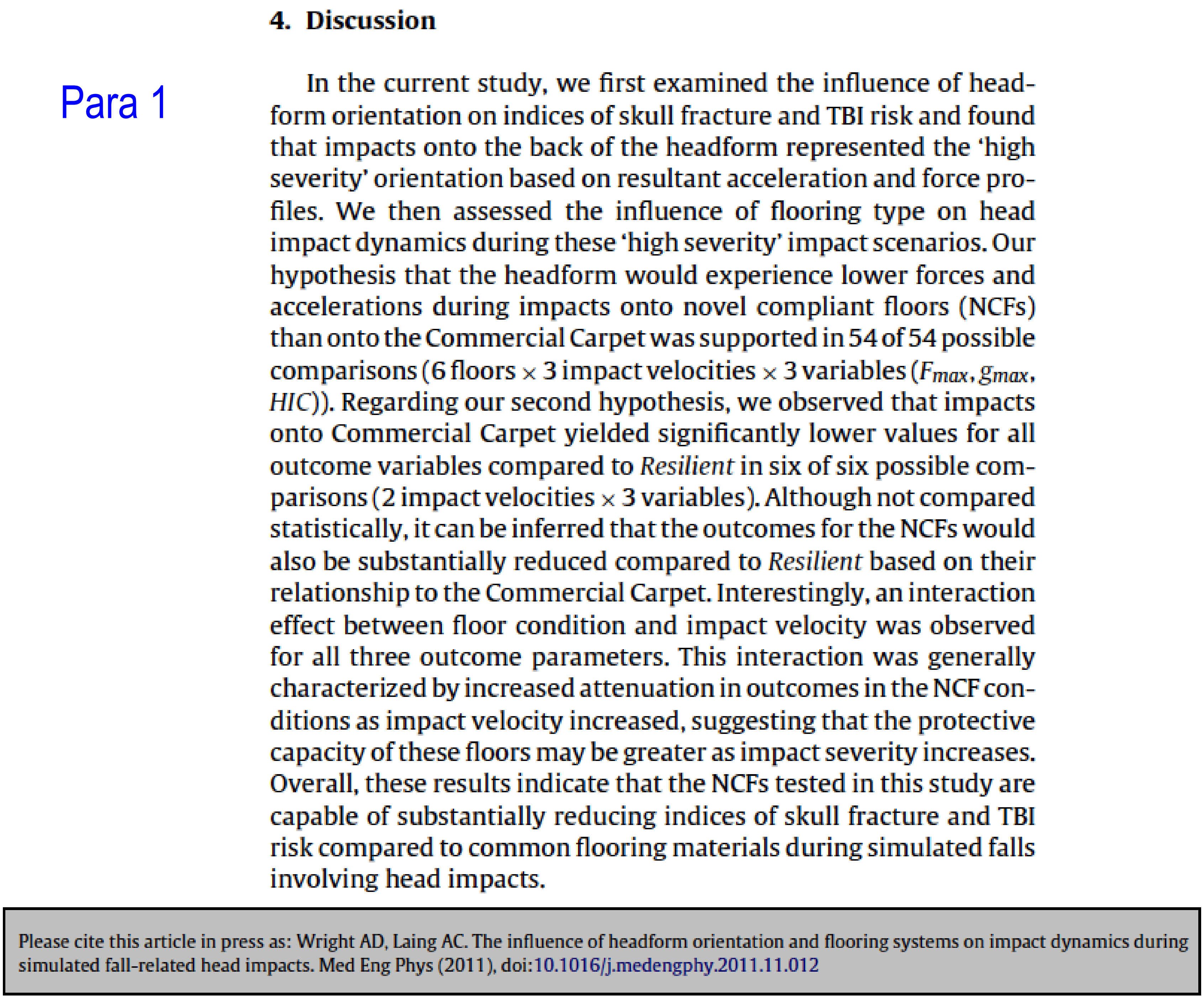
- Main purpose is to answer the question(s) posed in the Introduction.

- Funnel from specific to general.
  - Answer the research question/hypothesis by stating supporting evidence.
  - Explain how the answers fit with the existing knowledge on the topic.
  - The Discussion may also include:
    - indications of the newness and importance of the work
    - explanations of discrepancies with others' results
    - explanations of the limitations of the study
    - implications for clinical practice or future research

Project: Discussion, Wright & Laing 2011

- 1st para:
- 2nd para:
- 3rd para:
- 4th para:
- 5th para:
- 6th para:
- 7th para:
- 8th para:
- 9th para:

Para 1

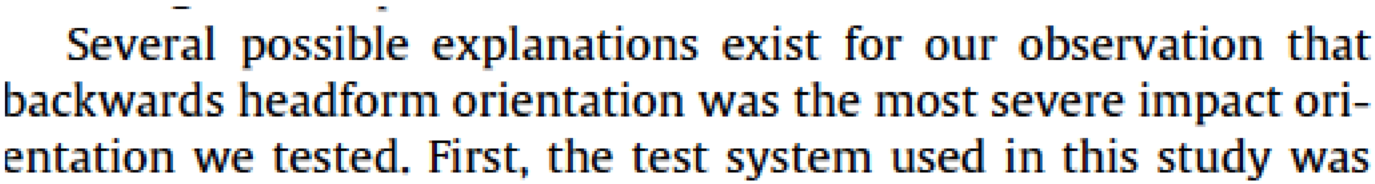


4. Discussion

In the current study, we first examined the influence of headform orientation on indices of skull fracture and TBI risk and found that impacts onto the back of the headform represented the ‘high severity’ orientation based on resultant acceleration and force profiles. We then assessed the influence of flooring type on head impact dynamics during these ‘high severity’ impact scenarios. Our hypothesis that the headform would experience lower forces and accelerations during impacts onto novel compliant floors (NCFs) than onto the Commercial Carpet was supported in 54 of 54 possible comparisons (6 floors × 3 impact velocities × 3 variables (F_{max}, g_{max}, *HIC*)). Regarding our second hypothesis, we observed that impacts onto Commercial Carpet yielded significantly lower values for all outcome variables compared to *Resilient* in six of six possible comparisons (2 impact velocities × 3 variables). Although not compared statistically, it can be inferred that the outcomes for the NCFs would also be substantially reduced compared to *Resilient* based on their relationship to the Commercial Carpet. Interestingly, an interaction effect between floor condition and impact velocity was observed for all three outcome parameters. This interaction was generally characterized by increased attenuation in outcomes in the NCF conditions as impact velocity increased, suggesting that the protective capacity of these floors may be greater as impact severity increases. Overall, these results indicate that the NCFs tested in this study are capable of substantially reducing indices of skull fracture and TBI risk compared to common flooring materials during simulated falls involving head impacts.

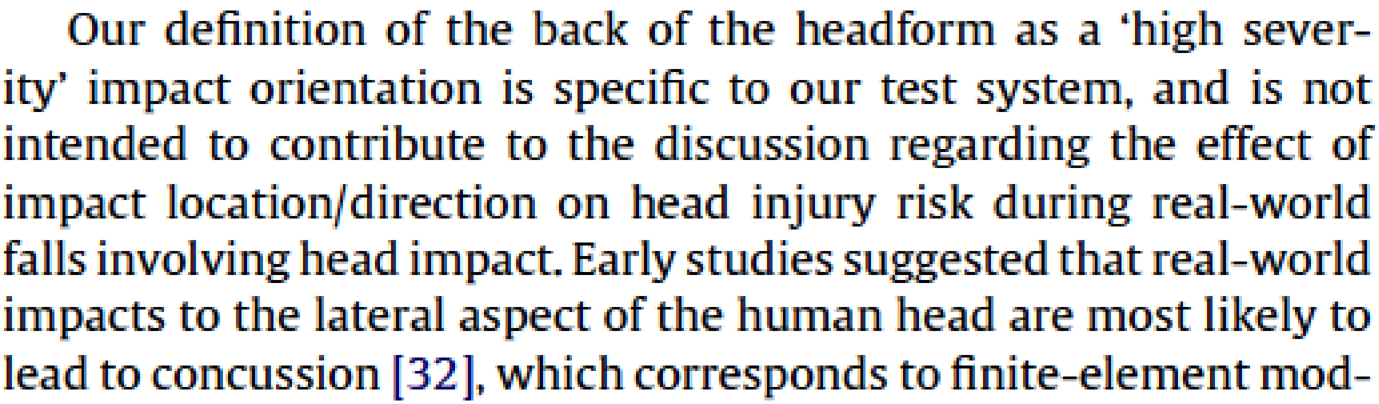
Please cite this article in press as: Wright AD, Laing AC. The influence of headform orientation and flooring systems on impact dynamics during simulated fall-related head impacts. Med Eng Phys (2011), doi:10.1016/j.medengphy.2011.11.012

Para 2



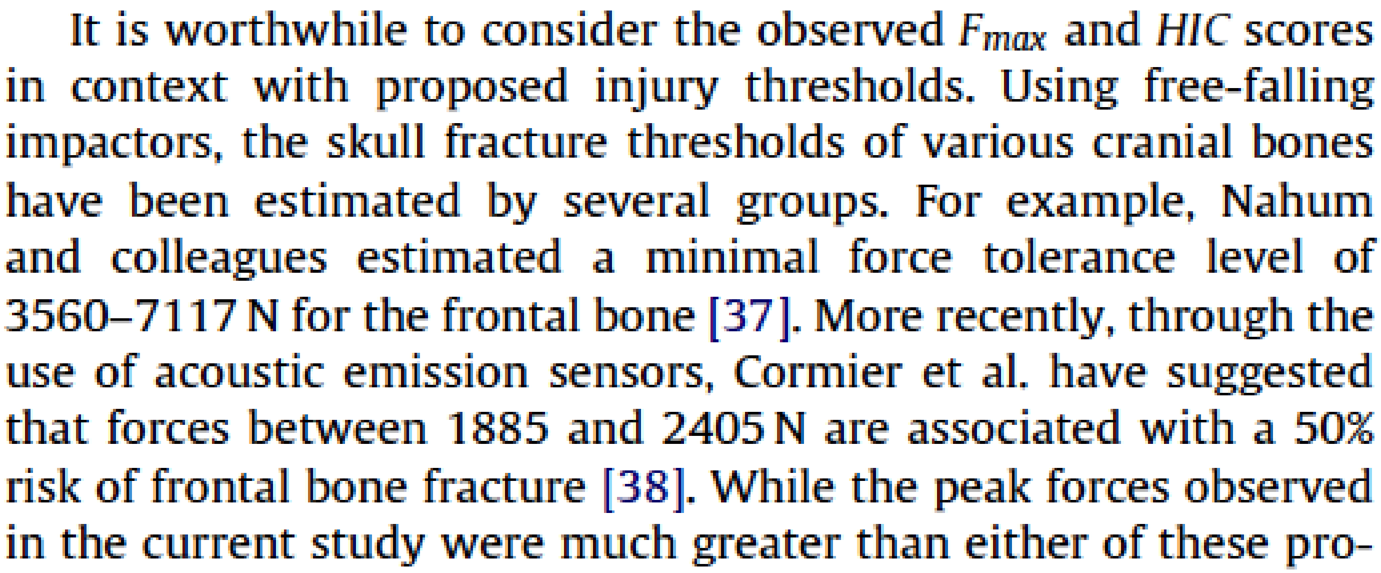
Several possible explanations exist for our observation that backwards headform orientation was the most severe impact orientation we tested. First, the test system used in this study was

Para 3

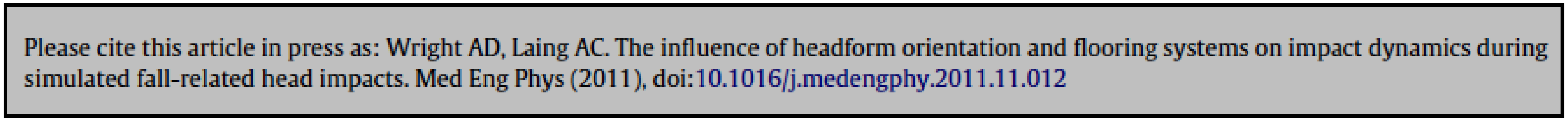


Our definition of the back of the headform as a 'high severity' impact orientation is specific to our test system, and is not intended to contribute to the discussion regarding the effect of impact location/direction on head injury risk during real-world falls involving head impact. Early studies suggested that real-world impacts to the lateral aspect of the human head are most likely to lead to concussion [32], which corresponds to finite-element mod-

Para 4

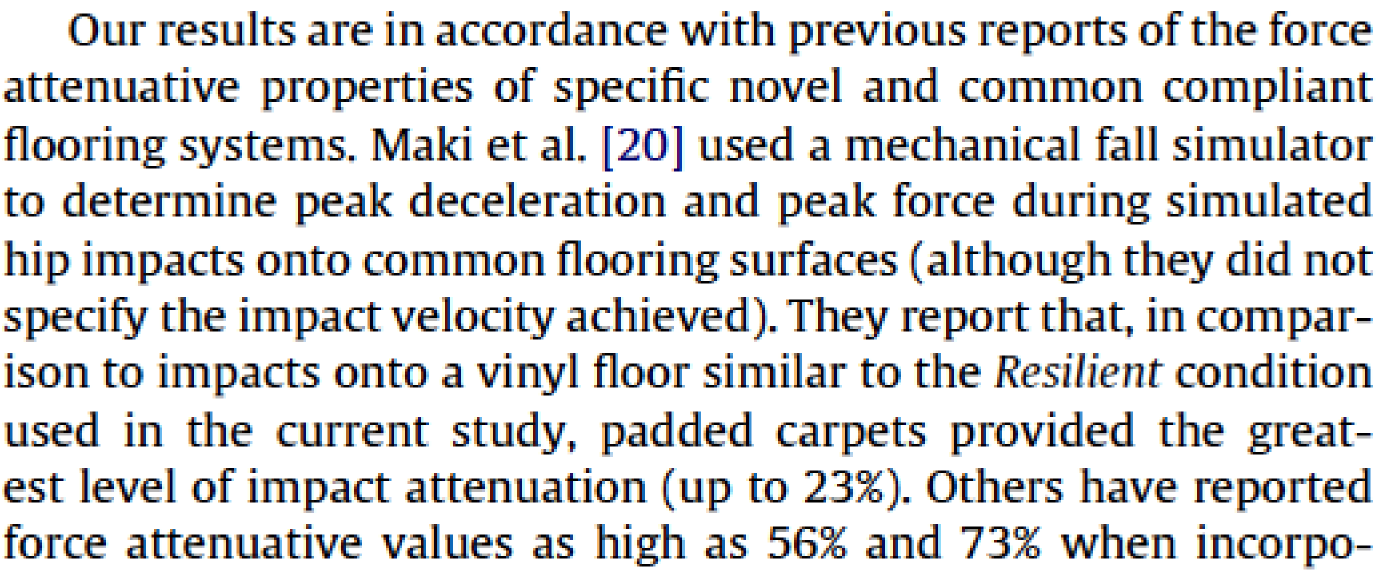


It is worthwhile to consider the observed F_{max} and *HIC* scores in context with proposed injury thresholds. Using free-falling impactors, the skull fracture thresholds of various cranial bones have been estimated by several groups. For example, Nahum and colleagues estimated a minimal force tolerance level of 3560–7117 N for the frontal bone [37]. More recently, through the use of acoustic emission sensors, Cormier et al. have suggested that forces between 1885 and 2405 N are associated with a 50% risk of frontal bone fracture [38]. While the peak forces observed in the current study were much greater than either of these pro-



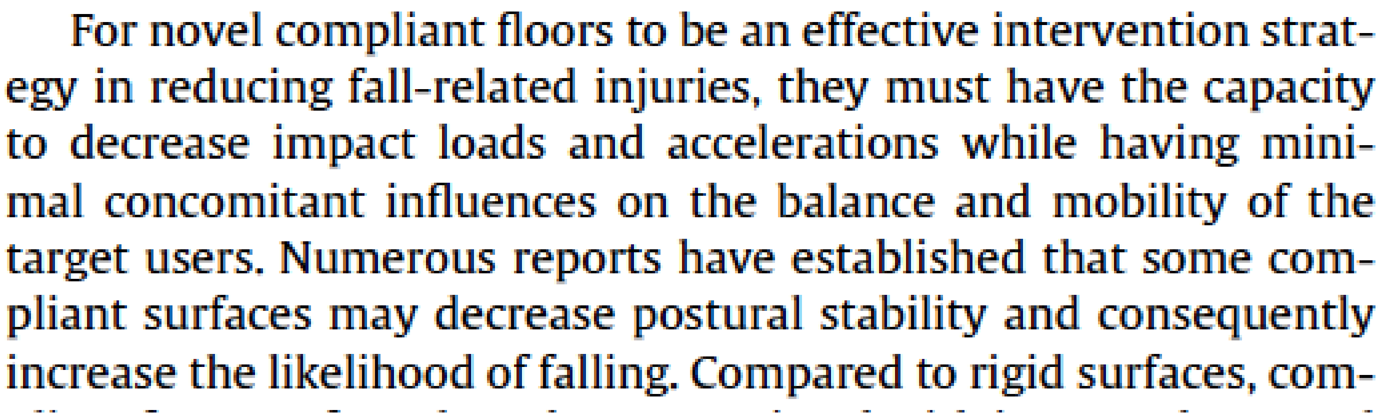
Please cite this article in press as: Wright AD, Laing AC. The influence of headform orientation and flooring systems on impact dynamics during simulated fall-related head impacts. Med Eng Phys (2011), doi:10.1016/j.medengphy.2011.11.012

Para 5

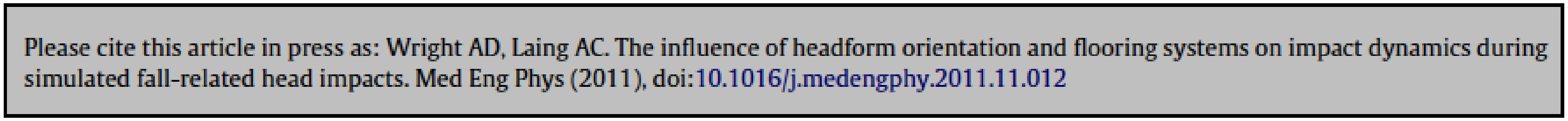


Our results are in accordance with previous reports of the force attenuative properties of specific novel and common compliant flooring systems. Maki et al. [20] used a mechanical fall simulator to determine peak deceleration and peak force during simulated hip impacts onto common flooring surfaces (although they did not specify the impact velocity achieved). They report that, in comparison to impacts onto a vinyl floor similar to the *Resilient* condition used in the current study, padded carpets provided the greatest level of impact attenuation (up to 23%). Others have reported force attenuative values as high as 56% and 73% when incorpo-

Para 6

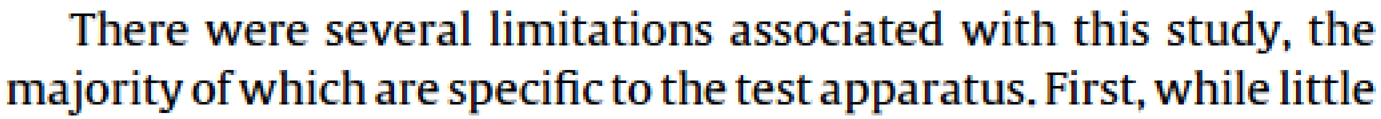


For novel compliant floors to be an effective intervention strategy in reducing fall-related injuries, they must have the capacity to decrease impact loads and accelerations while having minimal concomitant influences on the balance and mobility of the target users. Numerous reports have established that some compliant surfaces may decrease postural stability and consequently increase the likelihood of falling. Compared to rigid surfaces, com-



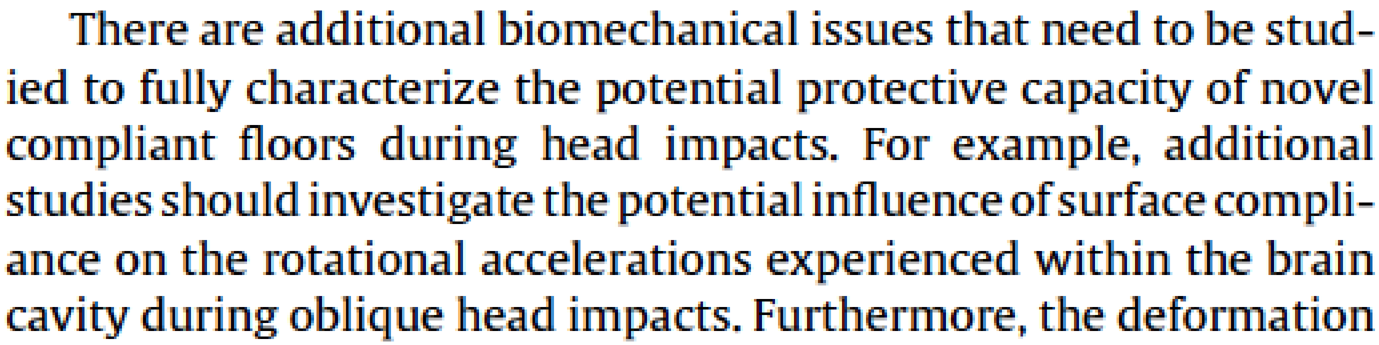
Please cite this article in press as: Wright AD, Laing AC. The influence of headform orientation and flooring systems on impact dynamics during simulated fall-related head impacts. Med Eng Phys (2011), doi:10.1016/j.medengphy.2011.11.012

Para 7



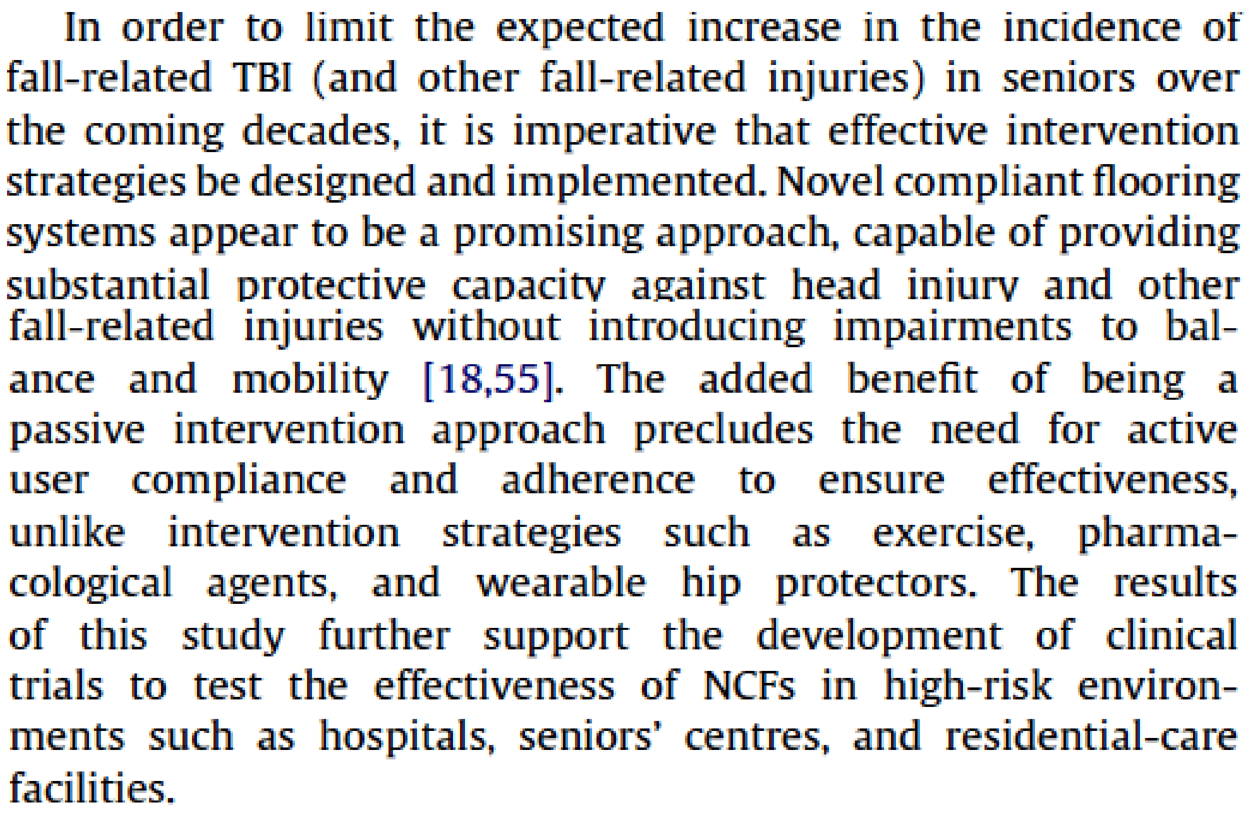
There were several limitations associated with this study, the majority of which are specific to the test apparatus. First, while little

Para 8



There are additional biomechanical issues that need to be studied to fully characterize the potential protective capacity of novel compliant floors during head impacts. For example, additional studies should investigate the potential influence of surface compliance on the rotational accelerations experienced within the brain cavity during oblique head impacts. Furthermore, the deformation

Para 9



In order to limit the expected increase in the incidence of fall-related TBI (and other fall-related injuries) in seniors over the coming decades, it is imperative that effective intervention strategies be designed and implemented. Novel compliant flooring systems appear to be a promising approach, capable of providing substantial protective capacity against head injury and other fall-related injuries without introducing impairments to balance and mobility [18,55]. The added benefit of being a passive intervention approach precludes the need for active user compliance and adherence to ensure effectiveness, unlike intervention strategies such as exercise, pharmacological agents, and wearable hip protectors. The results of this study further support the development of clinical trials to test the effectiveness of NCFs in high-risk environments such as hospitals, seniors' centres, and residential-care facilities.

Writer’s Corner
Active vs. Passive Voice

- Many people believe they should avoid the passive voice, but fewer people can define it or recognize it. So, let’s try to understand the difference between passive and active voices.
- In an active sentence, the subject is doing the action.
  - For example, "Steve loves Amy." Steve is the subject, and he is doing the action: he loves Amy, the object of the sentence.
  - For example, “I heard it during dinner.” “I” is the subject, the one who is doing the action. “I” heard “it”, the object of the sentence.

Grammar Girl, Podcast Episode 232: July 22, 2010 (www.grammar.quickanddirtytips.com)

Writer’s Corner
Active vs. Passive Voice

- In passive voice, the target of the action gets promoted to the subject position.
  - For example, instead of saying, "Steve loves Amy," I would say, "Amy is loved by Steve." The subject of the sentence is now Amy, but she isn't doing anything. She is simply the recipient of Steve's love. The focus of the sentence has changed from Steve to Amy.
  - For example, instead of saying, “I heard it during dinner,” I would say, “It was heard by me during dinner.”

- One clue that your sentence is passive is that the subject isn't taking a direct action

Grammar Girl, Podcast Episode 232: July 22, 2010 (www.grammar.quickanddirtytips.com)

Writer's Corner
Active vs. Passive Voice

- Is passive voice always wrong?
  - Passive sentences are not incorrect
  - But, passive voice is sometimes awkward, wordy, and vague
  - Using the active voice can tighten up your writing
  - Businesses and politicians sometimes use passive voice to intentionally obscure who is taking the action
    - For example, "Mistakes were made."
    - For example, "Your power will be shut off on Monday."
  - It is a good idea to stick to the active voice if you are writing for the general population.

Grammar Girl, Podcast Episode 232: July 22, 2010 (www.grammar.quickanddirtytips.com)

Writer’s Corner
Active vs. Passive Voice

- What about scientific writing in particular?
  - Scientists are often encouraged to use the passive voice as a way to increase the sense of objectivity in their writing
  - However, some scientific style guides allow for active voice.
    - For example, “We sequenced the DNA,” is active, while “The DNA was sequenced” is passive.
  - It is poor form for scientists to insert themselves into conclusions.
    - You wouldn’t say, “We believe this mutation causes cancer.”
    - However, you can still use the active voice in conclusions. For example, “The results suggest that this mutation causes cancer.”
- For more, see Grammar Girl Episode 231, July 21, 2010, “How to write clear sentences.”
- For more on the passive voice, see http://writingcenter.unc.edu/handouts/passive-voice/

Grammar Girl, Podcast Episode 232: July 22, 2010 (www.grammar.quickanddirtytips.com)

Nonparametric Statistics

- So far in this course, we've covered independent t-tests, paired t-tests, ANOVA, correlation, and linear regression.

- These parametric statistical tests require the dependent variable to be continuous and approximately normally distributed.

- But what if your data do not meet these criteria?

- Nonparametric statistical tests do not require the data to belong to any particular distribution. Therefore, nonparametric tests are appropriate if data are not continuous or normally distributed.

Nonparametric Tests

- **Is There a Difference?**
  - **Wilcoxson signed rank test:** Analogous to paired t-test.
  - **Wilcoxson rank sum test:** Analogous to independent t-test.
  - **Chi-square:** Analogous to ANOVA. It tests differences in the frequency of observations of categorical data.

- **Is there a Relationship?**
  - Rank Order Correlation: Analogous to the Pearson correlation coefficient . These tests examine relationships between ordinal variables. Includes the **Spearman's Rank Order Correlation** (rs) & **Kendall's Tau** (τ).

- **Can we predict?**
  - **Logistic Regression:** Analogous to linear regression. It assesses the ability of variables to predict a dichotomous variable.

Chi-square

- Imagine you’ve conducted a study of 50 men and 40 women. Of these 45 (90%) men were married and 38 (95%) women were married?
- In this example, marital status is a categorical dependent variable (married/not married). Since it is not continuous, you don’t compute the mean of marital status. Instead you compute the percentage of married men and women.
- This will lead to ask, “Were women more likely to be married than men?”
- You can use chi-square to determine whether the percentage of married women was significantly different from the percentage of married men.

Chi-square

- The chi-square tests for a difference in the **proportion of observed frequencies** across a given set of categories in comparison to the **proportion of expected frequencies**.

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

Simple Chi-square

# Right-handed	# Left-handed	Total
38	6	44

- In a study of 44 subjects we observed 6 left-handers and 38 right-handers
- If we are testing whether there are equal numbers of right and left-handers then the expected frequencies would be 22 and 22.
- Calculate the value of Chi-square:

Simple Chi-square

# Right-handed	# Left-handed	Total
38	6	44

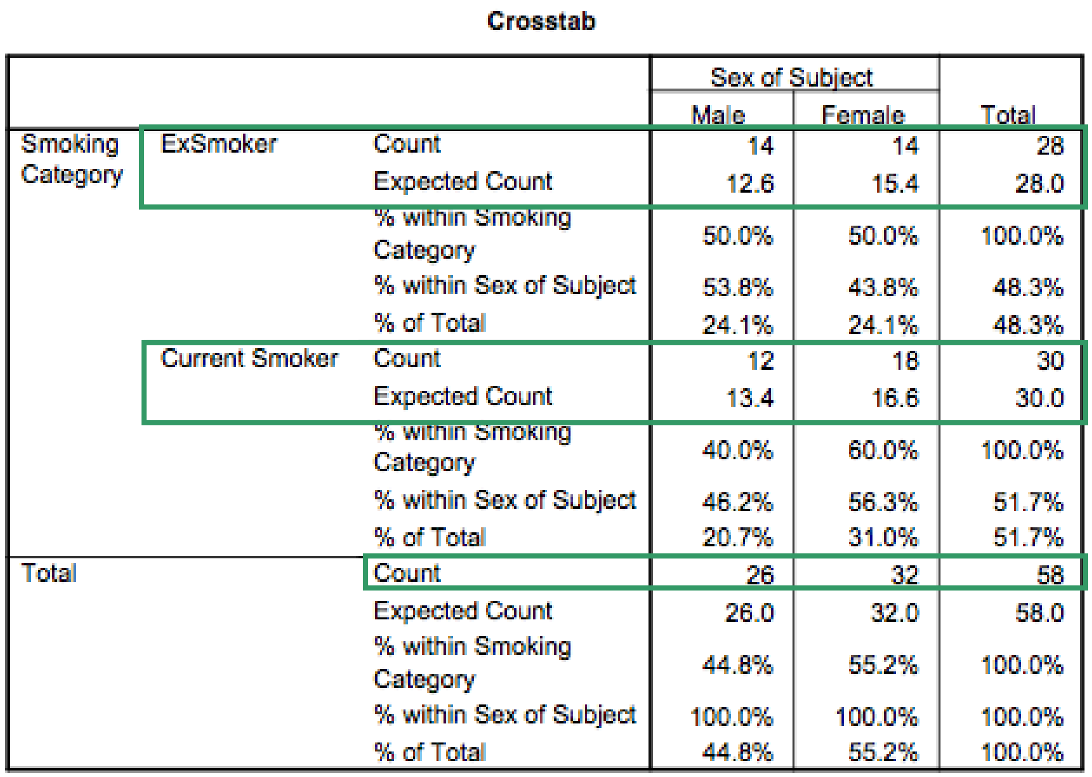
- In a study of 44 subjects we observed 6 left-handers and 38 right-handers
- If we are testing whether 15% of the sample is left-handed then the expected frequencies would be 6.6 (0.15 x 44) for left-handers and 37.4 (0.85 x 44) for right-handers.
- Calculate the value of Chi-square:

Two-way Chi-square

Is the distribution of smoking status different between men and women?

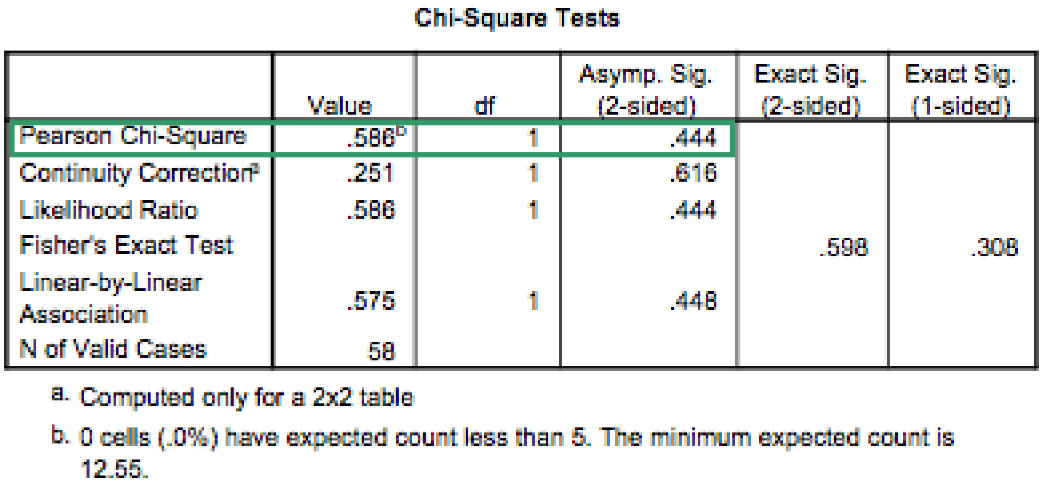
		Men (N=26)	**Women (N=32)**	**Total**
Ex-Smoker	Observed	14	14	**28**
	Expected			
Current smoker	Observed	12	18	**30**
	Expected			
	Total	**26**	**32**	**58**

Try to calculate the Chi-square value by hand. See text chapter.



Crosstab

			Sex of Subject		
			Male	Female	Total
Smoking Category	ExSmoker	Count	14	14	28
		Expected Count	12.6	15.4	28.0
		% within Smoking Category	50.0%	50.0%	100.0%
		% within Sex of Subject	53.8%	43.8%	48.3%
		% of Total	24.1%	24.1%	48.3%
	Current Smoker	Count	12	18	30
		Expected Count	13.4	16.6	30.0
		% within Smoking Category	40.0%	60.0%	100.0%
		% within Sex of Subject	46.2%	56.3%	51.7%
		% of Total	20.7%	31.0%	51.7%
Total		Count	26	32	58
		Expected Count	26.0	32.0	58.0
		% within Smoking Category	44.8%	55.2%	100.0%
		% within Sex of Subject	100.0%	100.0%	100.0%
		% of Total	44.8%	55.2%	100.0%



Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	.586[b]	1	.444		
Continuity Correction[a]	.251	1	.616		
Likelihood Ratio	.586	1	.444		
Fisher's Exact Test				.598	.308
Linear-by-Linear Association	.575	1	.448		
N of Valid Cases	58				

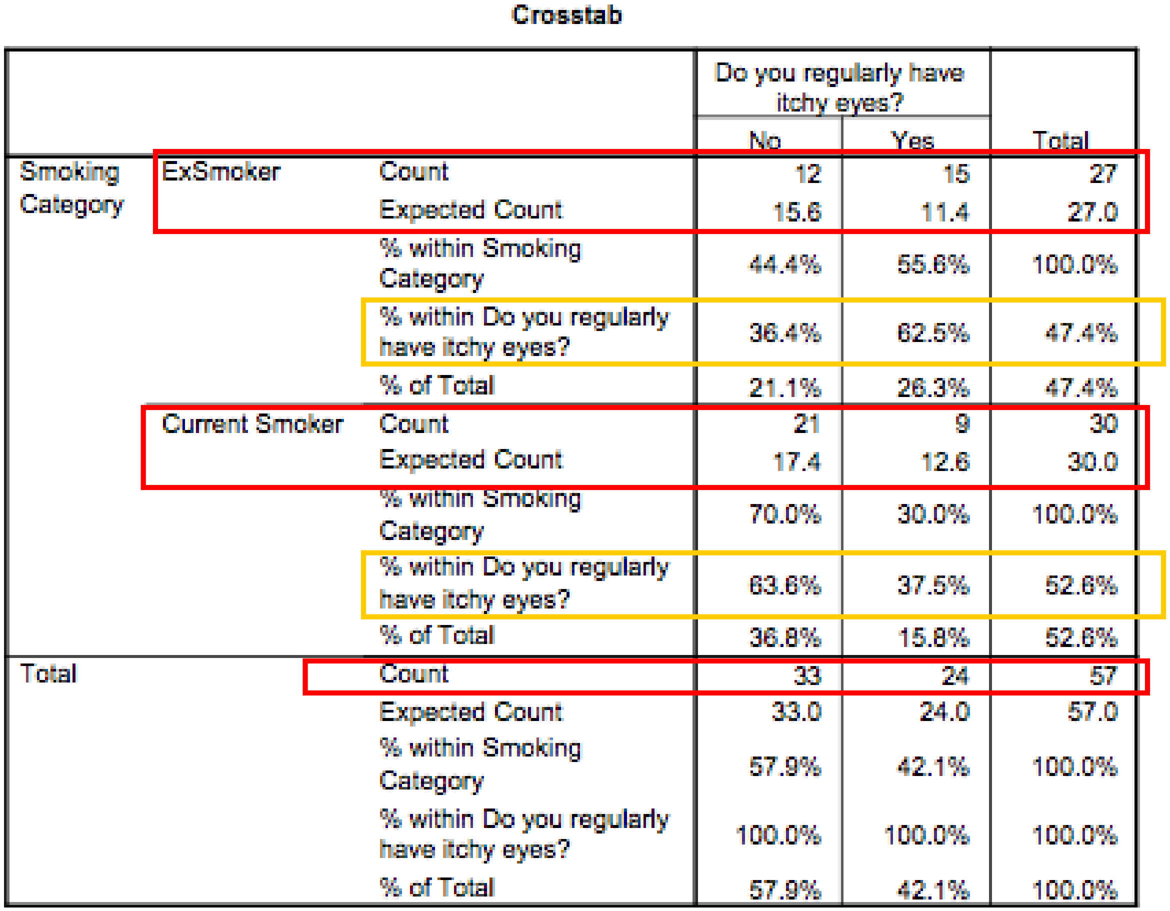
a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 12.55.

What do you conclude?

Do you regularly have itchy eyes? Yes or No?

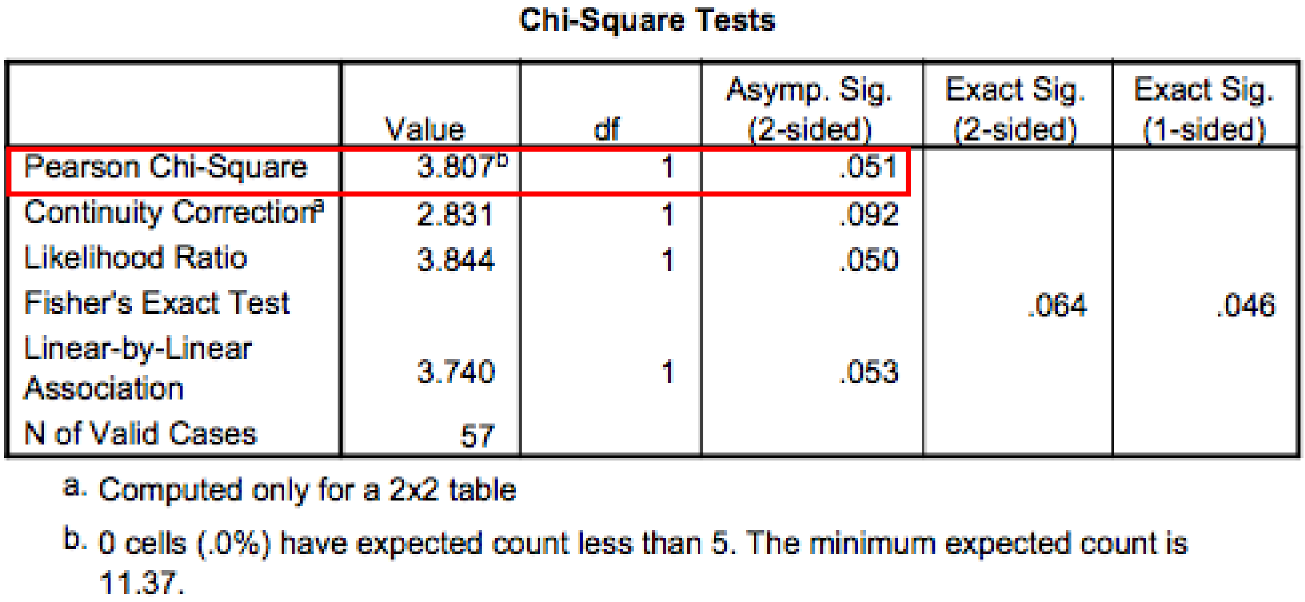
Is the distribution of smoking status different between those who do and do not report itchy eyes?



Crosstab

			Do you regularly have itchy eyes?		Total
			No	Yes	
Smoking Category	ExSmoker	Count	12	15	27
		Expected Count	15.6	11.4	27.0
		% within Smoking Category	44.4%	55.6%	100.0%
		% within Do you regularly have itchy eyes?	36.4%	62.5%	47.4%
		% of Total	21.1%	26.3%	47.4%
	Current Smoker	Count	21	9	30
		Expected Count	17.4	12.6	30.0
		% within Smoking Category	70.0%	30.0%	100.0%
		% within Do you regularly have itchy eyes?	63.6%	37.5%	52.6%
		% of Total	36.8%	15.8%	52.6%
Total		Count	33	24	57
		Expected Count	33.0	24.0	57.0
		% within Smoking Category	57.9%	42.1%	100.0%
		% within Do you regularly have itchy eyes?	100.0%	100.0%	100.0%
		% of Total	57.9%	42.1%	100.0%

“Do you regularly have itchy eyes? Yes or no?”



Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	3.807[b]	1	.051		
Continuity Correction[a]	2.831	1	.092		
Likelihood Ratio	3.844	1	.050		
Fisher's Exact Test				.064	.046
Linear-by-Linear Association	3.740	1	.053		
N of Valid Cases	57				

[a] Computed only for a 2x2 table

[b] 0 cells (.0%) have expected count less than 5. The minimum expected count is 11.37.

Conclusion: While it appeared that individuals who reported itchy eyes were more likely to be ex-smokers than individuals who did not report itchy eyes (62.5% vs. 36.4%), this result was not significant ($\chi^2(1)=3.807$, p=0.051).

OR

Conclusion: While it appeared that individuals who reported itchy eyes were less likely to be current smokers than individuals who did not report itchy eyes (37.5% vs. 63.6%) , this result was not significant ($\chi^2(1)=3.807$, p=0.051).

Summary of Two-way Chi-square

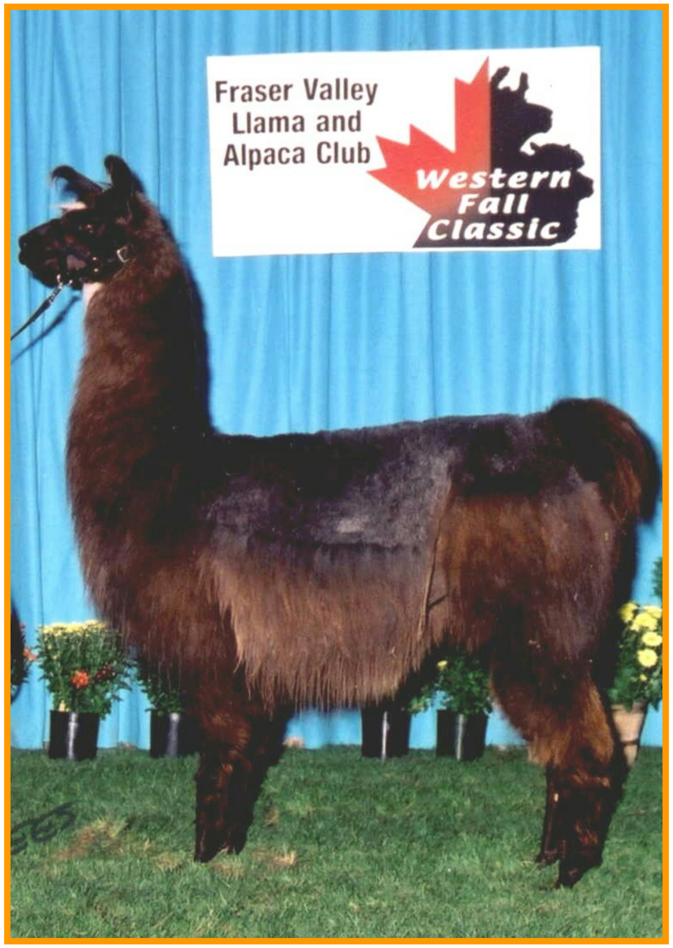
- Two categorical variables are considered simultaneously (e.g., sex and smoking status).
- Two-way Chi-square test is a test of independence between the two categorical variables.
- Null hypothesis: there is no difference in the frequency of observations for each variable in each cell.
- If the observed and expected frequencies are similar within each variable, the chi-square test will not be significant ($p \geq 0.05$).
- If the observed frequencies deviate considerably from the expected frequencies in one or more categories, the chi-square test will be significant ($p < 0.05$).

Spearman's Rank Order Correlation (r_s)

- You want to evaluate the relationship between variables, where neither of the variables is normally distributed.
- The calculation of the Pearson correlation coefficient (*r*) is not appropriate in this situation (if one of the variables is normally distributed you can still use *r*).
- If both are not normally distributed then you can use:
  - Spearman's Rank Order Correlation Coefficient (r_s)
  - Kendall's tau (τ).
  - These tests rely on the two variables being rankings.

Example of Spearman’s Rank Order Correlation (r_s)

Llama #	Judge 1	Judge 2	d	d^2
1	1	1	0	0
2	3	4	-1	1
3	4	2	2	4
4	5	6	-1	1
5	2	3	-1	1
6	6	5	1	1
			Σd	Σd^2
			0	8



Calculate r_s:

$$r_s = 1 - \frac{6\Sigma d^2}{n(n^2 - 1)}$$

Logistic Regression

- Logistic regression is analogous to linear regression analysis in that you produce an equation to predict a dependent variable from independent variables
- Linear regression used continuous dependent variables.
- Logistic regression uses categorical dependent variables.
- Most common to use binary dependent variables.
- Binary variables have two possible values
  - Yes or No answer to a question on a questionnaire
  - Had an event vs. did not have an event (e.g., cancer diagnosis)
- It is usual to code binary variables as 0 or 1 (e.g., no=0, yes=1)

Logistic Regression

- In a binary variable if coded with 1s and 0s, the mean of the binary variable will represent the proportion of 1s.
  - Sample size of 100
  - Hip fracture coded 1 and no hip fracture coded as 0
  - 80 events (hip fracture) and 20 non events (no hip fracture)
  - Mean of the variable Hip Fracture would be .80 which is also the proportion of hip fractures in the sample.
  - Proportion of not hip fracture would then be 1 – 0.8 = 0.2.
- The mean of the binary variable, and therefore the proportion of 1s, is labeled P
- The proportion of 0s is labeled Q, with Q = 1 - P
- In parametric statistics, the mean of a sample has an associated variance and standard deviation, so too does a binary variable.
  - Variance of a binary variable is PQ or P(1-P)
  - Standard deviation of a binary variable is $\sqrt{PQ}$ or $\sqrt{P(1-P)}$

Logistic Regression

- P not only tells you the proportion of 1s but it also gives you the probability of selecting a 1 from a random sample of the population.
  - 80% chance of selecting a participant with a hip fracture (in other words, the probability of hip fracture is 0.80)
  - 20% chance of selecting a participant without a hip fracture (in other words, the probability of no hip fracture is 0.20)

➡ Recall, probabilities range from 0 to 1.

Reasons why logistic regression should be used rather than ordinary linear regression in the prediction of binary variables

1. Predicted values of a binary variable cannot theoretically be greater than 1 or less than 0. This could happen however, when you predict the dependent variable using a linear regression equation.

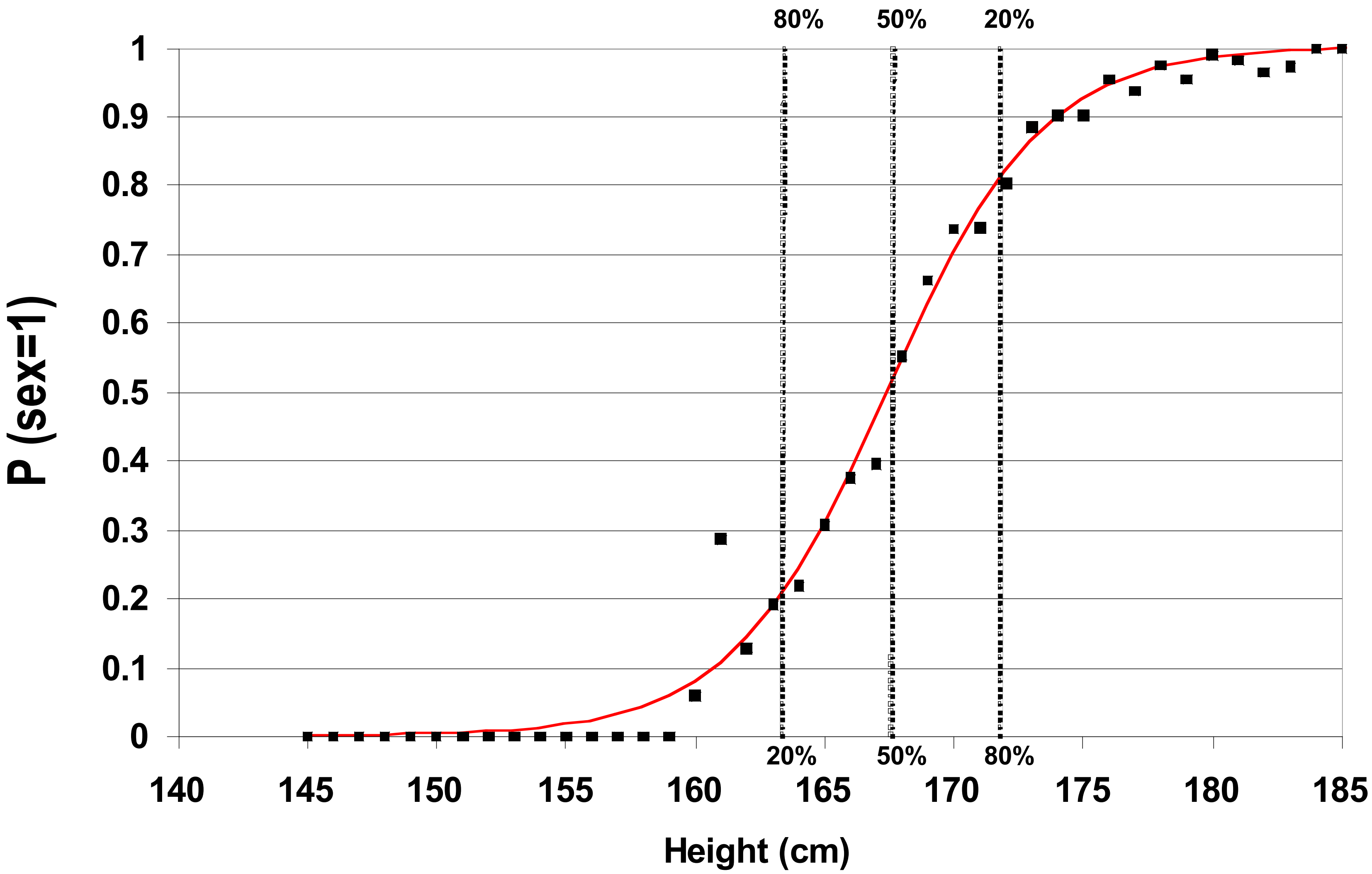
2. Linear regression requires that the residuals are normally distributed, but this is clearly not the case when the dependent variable can only have values of 1 or 0.

Reasons why logistic regression should be used rather than ordinary linear regression in the prediction of binary variables

3. It is assumed in linear regression that the variance of Y is constant across all values of X. This is referred to as homoscedasticity.
   - Variance of a binary variable Y is PQ. Therefore, the variance is dependent upon the proportion at any given value of the independent variable.
   - Variance of Y is greatest when 50% are 1s and 50% are 0s. Variance reduces to 0 as P reaches 1 or 0. This variability of variance is referred to as heteroscedasticity

P	Q	PQ Variance
0	1	0
.1	.9	.09
.2	.8	.16
.3	.7	.21
.4	.6	.24
.5	.5	.25
.6	.4	.24
.7	.3	.21
.8	.2	.16
.9	.1	.09
1	0	0

Canada Fitness Survey (1981): Logistic curve fitting through rolling means of binary variable sex (1=male, 0=female) versus height in cm



Odds & log Odds

Probability of being male at a height of 174 cm is .90. What are the odds and log odds of being male & female when height=174 cm?

Male $$Odds = \frac{P}{1-P} = \frac{0.9}{1-0.9} = 0.9/0.1 = 9$$

Female $$Odds = \frac{P}{1-P} = \frac{0.1}{1-0.1} = 0.1/0.9 = 0.11$$

The natural log of 9 is 2.217 [ln(.9/.1)=2.217]

The natural log of 1/9 is -2.217 [ln(.1/.9)=-2.217]

log odds of being male is exactly opposite to the log odds of being female.

Logit

In logistic regression, the dependent variable is a logit or log odds, which is defined as the natural log of the odds:

$$\operatorname{logit}(P) = \log(odds) = \ln\left(\frac{P}{1-P}\right)$$

In logistic regression, the estimated parameter is an Odds Ratio.

Odds Ratio

	Heart Attack	No Heart Attack	Probability of Heart Attack	Odds of Heart Attack
Treatment	3	6		
No Treatment	7	4		
			Odds Ratio:	

Recall that odds = P/(1-P)

Odds Ratio

	Heart Attack	No Heart Attack	Probability of Heart Attack	Odds of Heart Attack
Treatment	3	6	3/(3+6)=0.33	0.33/(1-0.33) = 0.50
No Treatment	7	4	7/(7+4)=0.64	0.64/(1-0.64) = 1.75
			Odds Ratio:	1.75/0.50 = **3.50**

The odds of a heart attack were 3.5 times greater among individuals who did not receive treatment compared to those who did receive treatment.

Linear vs. Logistic Regression Models

- General form of a linear regression model:

  $Y = B_1X_1 + B_2X_2 + B_3X_3 \ldots\ldots + B_o$

  Y is a continuous, normally distributed variable, e.g., blood pressure

- General form of a logistic regression model:

  $\text{Log odds } (Y) = B_1X_1 + B_2X_2 + B_3X_3 \ldots\ldots + B_o$

  Y is a binary variable, e.g., heart attack (yes/no)

You can predict probabilities from a logistic regression model

$$P = \frac{1}{1 + e^{-(B_0 + B_1 X)}}$$

- ***P*** is the probability of a 1 (the proportion of 1s, the mean of ***Y***)
- ***e*** is the base of the natural logarithm (about 2.718)
- ***B_0*** and ***B_1*** are coefficients from the logistic model.

Maximum Likelihood

- The loss function quantifies the goodness of fit of the equation to the data.
- Linear regression – least sum of squares
- Logistic regression is nonlinear. For logistic curve fitting and other nonlinear curves the method used is called maximum likelihood
  - Values for the coefficients (e.g., $\boldsymbol{B_0}$ and $\boldsymbol{B_1}$*)* are picked randomly and then the likelihood of the data given those values of the parameters is calculated.
  - Each one of these changes is called an iteration
  - The process continues iteration after iteration until the largest possible value or Maximum Likelihood has been found.

Allergy Questionnaire

Research Question: Are you more likely to have a cat allergy if your Mom or your Dad has a cat allergy?

catalrgy: Do you have an allergy to cats (No = 0, Yes = 1)
mumalrgy: Does your mother have an allergy to cats (No = 0, Yes = 1)
dadalrgy: Does your father have an allergy to cats (No = 0, Yes = 1)

Logistic Regression:

Dependent: catalrgy
Predictors: mumalrgy & dadalrgy

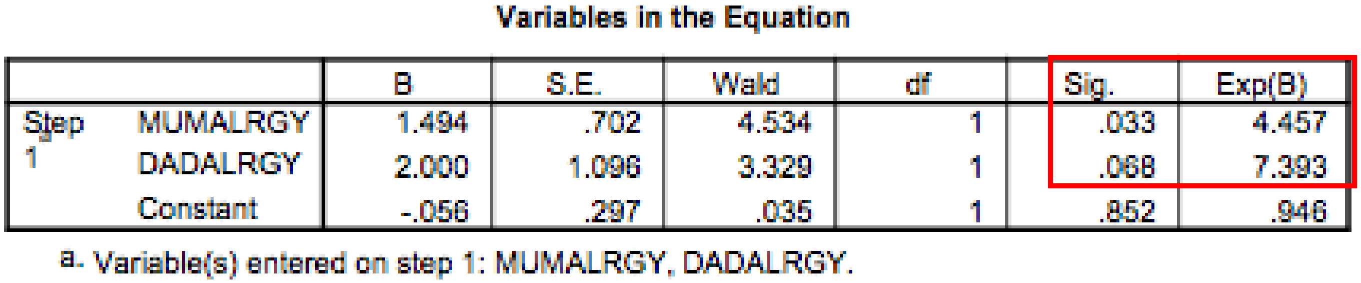
SPSS - Logistic Regression

Dependent: catalrgy

Predictors: mumalrgy & dadalrgy

Exp(B) is the Odds Ratio

If your mother has a cat allergy, your odds of having a cat allergy are 4.5 times higher than a person whose mother does not have a cat allergy (p=0.033).



Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	MUMALRGY	1.494	.702	4.534	1	.033	4.457
	DADALRGY	2.000	1.096	3.329	1	.068	7.393
	Constant	-.056	.297	.035	1	.852	.946

[a] Variable(s) entered on step 1: MUMALRGY, DADALRGY.

Log odds (CATALRGY) = -0.056 + 1.494(MUMALRGY) + 2.000(DADALRGY)