

Making Sense of Sensitivity: Using Candy and Anthropometric Data to Visually and Manipulatively Illustrate Sensitivity, Positive Predictive Value, and Related Terms

Brooke K. Bowman,^a Jason L. Furrer,^b Hannah C. Hart,^c Emily R. Wescott,^c and  Mark A. Milanick^a

^aDepartment of Medical Pharmacology and Physiology, University of Missouri, Columbia, Missouri, USA

^bDepartment of Microbiology and Molecular Immunology, University of Missouri, Columbia, Missouri, USA

^cDepartment of Nutrition and Exercise Physiology, University of Missouri, Columbia, Missouri, USA

KEYWORDS sensitivity, specificity, receiver operator curve, tactile, active, manipulative, screening, kinesthetic, negative predictive value, positive predictive value

INTRODUCTION

Students interested in a career in health care need to understand principles underlying diagnostic and screening tests, such as sensitivity and specificity, as they will communicate results to patients. While the concepts of sensitivity and specificity are useful to compare different tests, sensitivity and specificity by themselves cannot completely be used to determine test probabilities. Health care students can help by conveying to the patient the probability of having the condition using calculations based on sensitivity and specificity. All of these calculations are used when considering the tradeoffs of screening most or all of a population versus screening only a select population. Concepts of sensitivity and specificity are now front-and-center in the debates about screening for COVID. These concepts also have importance when screening for cancer, diabetes, and sexually transmitted infections. Thus, a better understanding of these concepts has importance to understanding, interpretation, and correct communications for health care workers, policy makers, and patients. Thus, not only future health care providers but also future health policy makers would benefit from a thorough understanding of these calculations.

The specific aims of this activity are to (i) improve teaching and understanding of sensitivity and specificity concepts, (ii) link sensitivity and specificity to prevalence, positive predictive value, and negative predictive value with a real-world context, and (iii) apply concepts from i and ii to make use of receiver operator

curves for determining the optimal cutoff for a particular test or to compare between tests.

Our working definitions of sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) are provided in Table 1. Our learning objectives were as follows: understand the calculation and concepts of sensitivity, specificity, PPV, and NPV, understand why sensitivity and specificity do not depend upon prevalence but PPV and NPV do, understand the value of reporting all 4 numbers, understand how to determine the optimal cutoff, and understand how to use area under the curve to determine the better test. While methodology of teaching sensitivity and specificity is discussed in several scholarly articles (1–4), the learners' understanding of sensitivity, specificity, and related concepts seems problematic in undergraduate curricula. At our institution, we could find only one life science course that covered these concepts. This prenursing course only discussed the terms as basic vocabulary. Of course, these concepts are covered in many statistical courses, but these tend to emphasize the mathematical aspects. The mathematical aspects appear disconnected from our life science students' everyday experience and use. We did a brief survey of first-semester first-year medical students and found that 5/8 had only been given the definitions in any undergraduate course and the other 3/8 had the material covered in part of a lecture. None confidently remembered what the terms meant after the class was completed. Zero said they saw or made any relevant connections to where that material would be used in a "medical future." After doing the activity we described here, we surveyed our students about their confidence of sensitivity and specificity concepts. Our anecdotal survey found that the students acknowledged a new understanding of the concept and appreciated the real-world significance of these terms after completing this exercise. The exercise also involved performing the relevant calculations, so they are likely to retain that understanding in the future.

In developing this activity, our goal was to create an exercise that actively engages the students and provides a more solid framework to build on compared to their previously

Editor Laura J. MacDonald, Hendrix College
Address correspondence to Department of Medical Pharmacology and Physiology, University of Missouri, Columbia, Missouri, USA.
E-mail: milanickm@missouri.edu.

The authors declare no conflict of interest.

Received: 22 October 2021, Accepted: 16 March 2022,

Published: 5 May 2022

TABLE I
Definitions

Question	Term
How well does the test detect people with the condition?	Sensitivity
How well does the test determine people that do not have the condition?	Specificity
What are the odds if one has a positive test that one has the condition?	PPV
What are the odds if one has a negative test that one does not have the condition?	NPV

oversimplified or superficial understanding of these concepts. In order to better teach sensitivity and specificity, we wanted kinesthetic and manipulative hands-on activities. These types of activities have the potential to improve the understanding of complex biological processes (5–8). We did a literature search on the Education Resources Information Center (ERIC) using the terms specificity and sensitivity. Only one slightly relevant primary article (9) was found. In this article, the hands-on approach involved performing Western blotting. The article discussed how sensitive and specific the different stains are. In spite of that, the article does not quantify these terms or directly relate the terms to clinical use. We identified an excellent case study (10) on teaching sensitivity and specificity. However, it takes about 5 class

periods and is math-intensive at a level that may be off-putting to undergraduates. Our activity can be done in one or two class periods and is more approachable for math skills. Our activity involves everyday objects and is visual and kinesthetic. We feel that our activity would have broader appeal. It would be particularly appropriate for lower-level undergraduate classes with students lacking solid math or graph-reading skills. We feel that the case study (10) would be a great follow-up to our activity or repeated later in students' upperclassman years.

An innovative way to accomplish teaching about sensitivity and specificity is to frame it with simple, manipulatable objects available to almost any teaching modality, and thus, we used candy and marbles. We motivated the activity by framing the learning into real-world scenarios such as the following. "Imagine someone has suggested that all students at a university should be tested for sexually transmitted infections (STIs) in order to reduce the stigma for getting an STI test. Another person argues that only those at high risk should be tested. How would calculations of sensitivity and specificity help here?" We explain to the students that by sorting and using the candy to represent patients in the scenario, they will learn about some of the tradeoffs of whether it is more important to reduce false positives or false negatives.

There are two parts to the activity. In the first part, a health care-related scenario was presented involving screening a whole population or a subset. Then, the students were given an array of candy (see Fig. 1). The candies represent the patients. The patients can be classified as those that have the disease and

The upper left group are red M&Ms
The upper right group are red skittles
The lower left group are not red M&Ms
The lower right group are not red skittles
Sensitivity

Red is a positive test,
not red is a negative test.
M&M have the condition,
skittles do not have the condition.
Specificity

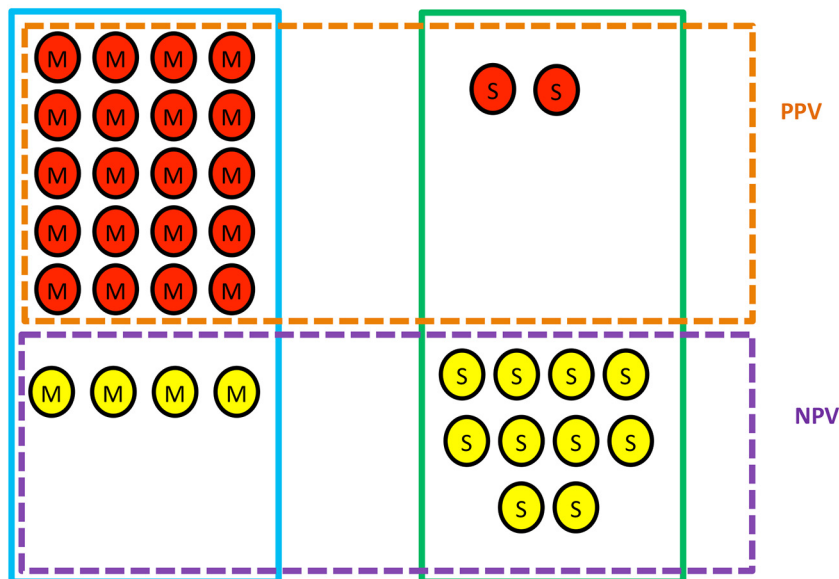


FIG 1. Using M&Ms and Skittles to explain screening test characteristics. In this format, the first column is used to calculate the sensitivity: How many of the M&Ms are red? The second column is used to calculate the specificity: How many of the Skittles are not red (yellow)? The top row is used to calculate the positive predictive value: If red, what are the odds it's an M&M? The bottom row is used to calculate the negative predictive value: If yellow (not red), what are the odds it's a Skittle?

those that do not have the disease. The patients can also be classified as those that have a positive test and those that have a negative test. Thus, there are 4 types of patients. These 4 types of patients are represented by the different properties of the candy. We guided the students through the calculation and terminology of the 4 parameters for this distribution of candy. Then, they examined different distributions to learn the effect of a change in prevalence (screening a whole population versus those with symptoms) on the calculations.

In the second component of the activity, we tested students' understanding of sensitivity and specificity concepts by having them apply calculations to receiver operator curves (ROCs). ROCs are plots of sensitivity versus $(1 - \text{specificity})$ and are often used to determine the optimal cutoff between a positive test and a negative test. ROCs can also be used to compare between different tests for the same condition in order to decide which surrogate test is better to use, compared to a gold standard that is invasive, expensive, and/or inconvenient. ROCs can seem daunting to students, so our activity using real-world data also motivates students and gets them to apply concepts. An example for this combinatorial learning is the following. "Percent body fat can be measured very accurately by hydrostatic weighing (inconvenient), dual-energy X-ray absorptiometry (expensive and radiation exposure), and BOD POD (expensive). As alternatives, researchers have used ROCs to evaluate different surrogate measurements, including neck circumference, waist circumference, waist hip ratio, and skin fold calipers. We asked the students to calculate ROCs of simple human body measurements to determine the optimal cutoff value to distinguish males from females. Can you determine which physical measurement is best to predict males versus females?" We feel that this lets students see the relevance to explanations they will make in their future careers and provides rationale for why specific testing or results were done.

We feel that these hands-on activities will lead to a more accurate understanding of the tradeoffs for screening and improve their future communication to patients or public for others.

PROCEDURE

Audience

This activity is flexible and can easily be done in 1 h. It can be done in a classroom setting, as a laboratory exercise, or as an outreach activity. Because of COVID restrictions, we also successfully tested this activity online, using the presentation and guiding slides provided in the appendices. We have done this activity in a variety of undergraduate classes, including Filtering Fact from Fiction in TV Crime and Medical Dramas (a general education class for all majors from freshman to seniors), Microbiology (for prenursing students, typically freshman or sophomores), and Translational Nutrition (for Nutrition and Exercise Physiology majors, typically juniors or seniors). In addition, it could easily be adapted for high school student levels or at any level in which the students have mastered fractions (if the students are also provided with more targeted guiding questions).

Overview of activity

The activity has 2 parts. For part 1, we have done this activity in two different versions. The goal of both versions was to teach students how to do calculations related to sensitivity and specificity and to discover the influence of prevalence on the parameters. The goal of part 2 was to have the students create a receiver operator curve and directly apply these concepts to a different set of real-world data.

In one version of part 1, the example test uses candy to represent patient conditions in testing. The result of testing is matched to whether the candy is red (positive) or not red (negative). The variable of having the condition is being an M&M and that of not having the condition is being a Skittle. We gave the students a set of candy, and the students had to calculate the sensitivity, specificity, PPV, and NPV in an easy-to-interpret grid format. Then, we had the students alter the prevalence of the condition and repeat the calculations to discover the influence of prevalence on the parameters. Initially, we used red and not-red objects, but the feedback from students suggested that having only two defined colors was easier to learn and more inclusive for color-blindness.

In the variation of part 1, a similar strategy is used but with the addition of using bags to conceal the physical properties (such as color in part 1) for kinesthetic sizing and having students be more predictive than analytical. In this version, the test is whether the object is large or small. The condition is being a marble and not having the condition is being a gumball. We gave the students 3 sets of bags, each set with different numbers of gumballs and marbles of two sizes. The students had to calculate the sensitivity, specificity, PPV, and NPV and decide which measurement would give the most favorable outcome if one wanted a gumball, which showed the influence of prevalence on the parameters.

In the supplementary materials, we have the slides that we used when presenting this online. Supplemental file 1 is for version 1 of the first half of the presentation, with M&Ms and Skittles, where the students do the calculations and then compare those values when the prevalence in the "population" changes. Supplemental file 2 is for version 2 of the first half of the presentation, where the students compare the sizes of gumballs and marbles. The students are given 3 "bags" with different prevalences and perform the calculations, compare data, and discuss implications for the 3 different conditions. Supplemental file 3 is for the second half, the activity examining the receiver operator curve.

MATERIALS AND PREPARATION

As this activity uses only candy and marbles, there are no safety issues.

For an in-person activity, one could provide each student with 20 red M&Ms, 2 red Skittles, 4 yellow M&Ms, and 10 yellow Skittles for version 1 or 20 small marbles, 2 small gumballs, 4 large marbles, and 10 large gumballs for version 2. A small

bowl, a small plastic box, or a plastic bag can be used to hold one set for each student. The students are instructed to place the items in the pattern shown in Fig. 1 on a level surface, or the instructor can have the arrangement done in advance to save time. We had each student work individually online. One could have in-person students work in pairs or triplets to prompt some small group discussion on the topics and let students teach each other.

The students explore how a change in prevalence would alter the results of the calculation. In the first case, prevalence is decreased but sensitivity is not changed. The students are given step-by-step instructions to guide them to complete a table to calculate the effect on the values in the case of decreasing prevalence (see supplementary file 1, pages 13 and 14 and supplementary file 2, especially pages 14 and 15). At this point, they discover that for a positive test, if the prevalence decreases, the odds of having a disease decrease. Then,

they are asked to complete a similar table for the situation when prevalence increases without step-by-step help.

Activity

To frame the activity and demonstrate the real-world usage of these concepts, a health care-related scenario is presented; examples are in slide one of supplementary material 1 and 2. The students then learn the concepts of specificity and sensitivity as well as PPV and NPV while performing one of the 2 variations of the hands-on activities (M&Ms/Skittles or marbles/gumballs). They then apply that knowledge and understanding to a new set of data, working with receiver operator curves (supplementary material 3).

After the groups complete either version 1 or version 2 of part 1, we have them examine how to determine the optimal cutoff and which surrogate test is better, as well as

TABLE 2
Selected feedback

Student comment
I really enjoyed how I was able to relate the M&M and skittles to the real-world scenario that we picked at the start. This was able to help me gather a better understanding of which of the areas were needing to be calculated based on the category they fell under.
I am a very big visual learner so it was really beneficial when the horizontal and vertical blocks started.
I loved the scenarios and the idea of discussing relatable tests and realistic illnesses that your students will deal with.
I think this would be a great addition to healthcare majors.
I liked the idea of the activity about heights because it's really interesting.
Big lightbulb moment right around this point
I really liked the visuals inside the presentation
I think the scenarios were easy to understand and relevant
I really enjoyed it and felt like I learned a lot from it.
Gumball and marble analogy is spot on. . . how do you even think of this?
The math was very manageable and not too hard
I developed a deeper understanding of this concept. I really liked the analogies presented. I also feel like I will not forget this concept due to I can relate it to very simple things like marbles and gumballs.
I like the m&m and skittle activity I thought the second activity was harder to understand because the first activity was easier to understand visually, in my opinion
I would not use "not red", I would stick to black or a different color. I also would use something more different, maybe like fruits and veggies
I would better explain what the activities are specifically about, preface the activity
To better improve this lab experiment, I would spend more time explaining what exactly this is about. Be straightforward, and concise with the word jargon. Explain to them why they are doing this, and what exactly it is about. Since I am a microbiology student, I know why I am doing this activity. But, I think for some people they need to be walked through the prevalence of it, as well as maybe before switching slides you explain what you will be showing next. It was very interesting, and I learned a lot!!
The only thing that I would suggest would be to explain more in detail. I felt like we went from cancer scenarios with candy to talking about the Army with no explanation. Other than that I thought it was great!
I really enjoyed how I was able to relate the M&M and skittles to the real-world scenario that we picked at the start. This was able to help me gather a better understanding of which of the areas were needing to be calculated based on the category they fell under. And I am a very big visual learner so it was really beneficial when the horizontal and vertical blocks started
I like how we were taught about the topics in simple forms of candy and the related it with true numbers of people in the military. The only annoying this is the numbers were difficult to understand on the real-life examples. This is something that I would need to be taught slower and maybe have the opportunity to do the calculations myself. Seeing the visual and doing the actual math helped me understand the topics we learned.

explain why they concluded that. We frame this with examples, including assessment of body fat and determination of insulin resistance (slides 1 and 2 of supplementary file 3). For the analysis the students will perform, we made use of the 2012 US Army Anthropometric Survey (ANSUR II) (11). We show them histograms of the height measurements for males and females in the US Army and use these data to create a receiver operator curve, which is a plot of sensitivity versus (1 – specificity). Using the ROC, the students can determine the optimal height cutoff as a surrogate measurement for whether the soldier is male or female. They can then compare other physical measurements to determine which is a better predictor of whether a soldier is male or female. While the better predictor has a larger area

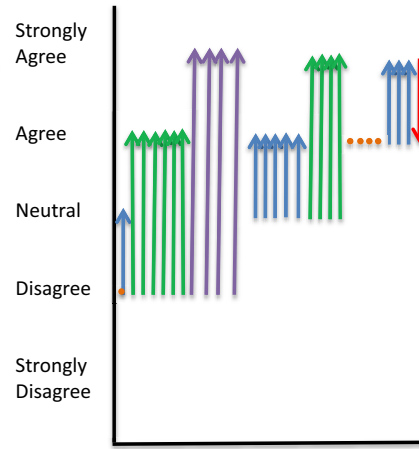
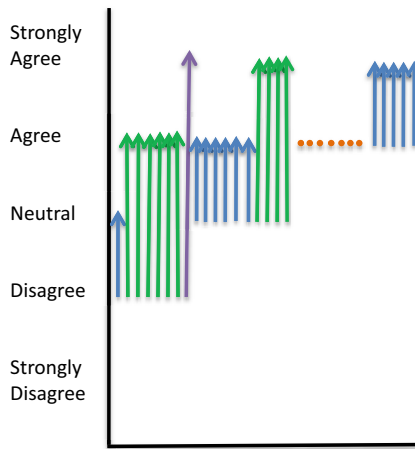
under the curve, it is important that the students understand why this is true. The closer the area under the curve is to 1, the better the test. However, it is important that the students note that the worst test has an area of 0.5, not zero.

DISCUSSION

We found that the authors (3 students and 2 instructors) learned new teaching techniques in designing this activity. For example, we now appreciate the use of an organized 2 by 2 layout to clearly note concepts to students. Also, we found that the concrete example of how prevalence alters the odds (when sensitivity and specificity are constant) provided us with a

Q1: I understand why many tests are described in terms of sensitivity and specificity even though what patients want to know is their odds of having the disease.

Q2: I have a solid understanding of sensitivity and specificity.



Q3: I understand how one determines the cut-off for surrogate diagnostic tests

Q4: I understand how one can use a receiver operator curve to compare 2 diagnostic tests.

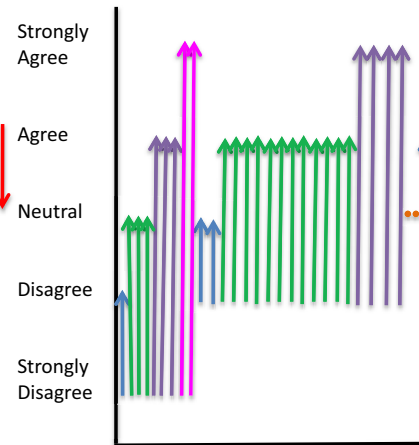
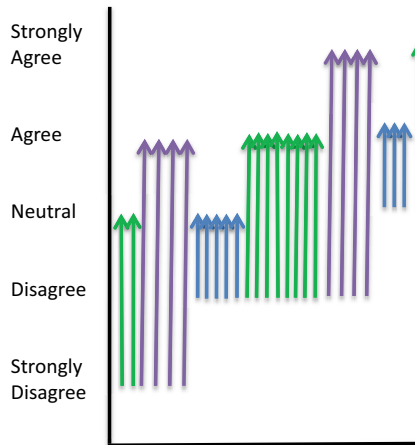


FIG 2. Feedback on the activity. For each question, the student was asked to rate their agreement on a Likert scale. Each line connects the before and after response of one student; the blunt end of the line is the student’s response before the activity and the arrowhead end of the line is the response after the activity. If there is no line, just a data point, the response did not change. Blue lines indicate an increase of one category, green lines an increase of 2 categories, purple lines an increase of 3 categories and red lines an increase of 4 categories. The black line indicates a decrease of 1 category.

greater appreciation of why this is true. Stressing this observation to students no doubt had the same impact on them. This applied both for those of us who are less mathematically inclined and for those that have a strong math background. Finally, in working out the activity for receiver operator curves, we gained a better appreciation of the source of frustration for students generating and analyzing these curves. Hopefully, we have found and present here interesting and thought-provoking ways to help keep students engaged and thinking critically about these data, the plots, and the presentation of, as well as the interpretations and conclusions that are drawn from, such plots.

One advantage of the second version of part I (marbles and gumballs) is that one could imagine having a dark bag containing both gumballs and marbles. If one just reached in, they could see how well using size to judge the difference between gumballs and marbles would work to predict which object is a gumball. Can the students apply the parameters they just calculated to a predictive scenario? This is analogous to a real-life test where the gold standard involves a biopsy (looking inside the bag) and one can ask how well a surrogate test (examining size) works.

Several students provided useful feedback on the value of the relevant scenarios to engage them at the beginning. Some of the student comments are found in Table 2; the negative student comments were mostly about the lack of motivation for the activity and the fact that for some class sections, we rushed through the activity. Other negative comments related to the confusion about the objects and how to do the calculations. The version in the appendices corrects these issues. The University of Missouri IRB determined that this activity is a Quality Improvement Project not requiring MU HRPP/IRB review.

We did a pretest and a posttest to evaluate the students' perception of the potential value of this activity and the presented concepts. The results are shown in Fig. 2. In addition to noting that almost every student felt that their understanding had increased, one can also note that the pretest degree of understanding of nearly every student was poor, indicating the need for covering this material at some point in the curriculum, given the personal and public policy decisions that can be affected by sensitivity and specificity of different diagnostic and screening tests.

Conclusion

We found that the visual and kinesthetic aspects helped the students scaffold and that starting with relevant use scenarios captured the attention of students. Using familiar objects engaged the students and made interpretation and understanding easier. Comparing the use of surrogate tests and prediction of physical characteristics to distinguish groups made the receiver operator curve portion more concrete for the students. Overall, we found that the students were engaged in the activity and found it useful.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

SUPPLEMENTAL FILE 1, PDF file, 2.7 MB.

SUPPLEMENTAL FILE 2, PDF file, 2.7 MB.

SUPPLEMENTAL FILE 3, PDF file, 7.7 MB.

ACKNOWLEDGMENTS

We report no funding. We do not have any conflicts of interest to declare.

REFERENCES

1. Baratloo A, Hosseini M, Negida A, El Ashal G. 2015. Part I: simple definition and calculation of accuracy, sensitivity and specificity. *Emerg (Tehran)* 3:48–49.
2. Loong TW. 2003. Understanding sensitivity and specificity with the right side of the brain. *BMJ* 327:716–719. <https://doi.org/10.1136/bmj.327.7417.716>.
3. Safari S, Baratloo A, Elfil M, Negida A. 2015. Evidence based emergency medicine part 2: positive and negative predictive values of diagnostic tests. *Emerg (Tehran)* 3:87–88.
4. Trevethan R. 2017. Sensitivity, specificity, and predictive values: foundations, plabilities, and pitfalls in research and practice. *Front Public Health* 5:307. <https://doi.org/10.3389/fpubh.2017.00307>.
5. Boomer SM, Latham KL. 2011. Manipulatives-based laboratory for majors biology - a hands-on approach to understanding respiration and photosynthesis. *J Microbiol Biol Educ* 12:127–134. <https://doi.org/10.1128/jmbe.v12i2.245>.
6. Cherney ID. 2008. The effects of active learning on students' memories for course content. *Active Learning in Higher Education* 9:152–171. <https://doi.org/10.1177/1469787408090841>.
7. Debruyen JM. 2012. Teaching the central dogma of molecular biology using jewelry. *J Microbiol Biol Educ* 13:62–64. <https://doi.org/10.1128/jmbe.v13i1.356>.
8. Guzman K, Bartlett J. 2012. Using simple manipulatives to improve student comprehension of a complex biological process: protein synthesis. *Biochem Mol Biol Educ* 40:320–327. <https://doi.org/10.1002/bmb.20638>.
9. Chang M-M, Lovett J. 2011. A laboratory exercise illustrating the sensitivity and specificity of Western blot analysis. *Biochem Mol Biol Educ* 39:291–297. <https://doi.org/10.1002/bmb.20501>.
10. COMAP, Inc. Imperfect testing: breast cancer case study. COMAP, Inc., Bedford, Massachusetts. https://www.comap.com/undergraduate/projects/biomath/PDF/Imperfect_Testing_SE.pdf.
11. Gordon CC, Blackwell CL, Bradtmilller B, Parham JL, Barrientos P, Paquette SP, Corner BD, Carson JM, Venezia JC, Rockwell BM, Mucher M, Kristensen S. 2012. Anthropometric survey of U.S. army personnel: methods and summary statistics. NATICK/TR-15/007. US Army Natick Soldier RD&E Center, Natick, MA. <https://www.hsdl.org/?view&did=762624>.