

CSE711

Topics in Differential Privacy

Marco Gaboardi

Room: 338-B

gaboardi@buffalo.edu

<http://www.buffalo.edu/~gaboardi>

Data



Aol.



NETFLIX

Google

Data



Aol.

Statistics over Data



NETFLIX

Google

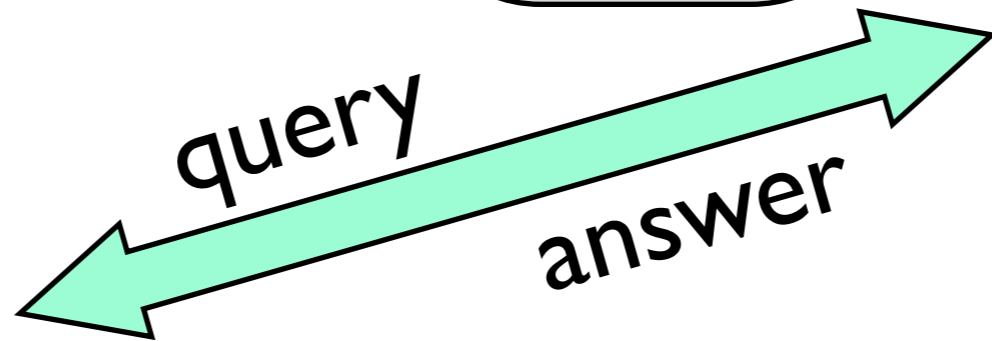
Private Queries?



Private Queries?



medical correlation?



Letter to the Editor

Methodological weakness in using correlation coefficients for assessing the interchangeability of analyte data between samples collected under different sampling conditions – the example of matrix metalloproteinase 9 determined in serum and plasma samples

Kuan Sheng¹ and Chien-Ying Ho²

¹Department of Forensic Chemistry, National Sun Yat-sen University, Kaohsiung, Taiwan; ²Institute of Forensic Chemistry, National Sun Yat-sen University, Kaohsiung, Taiwan

¹Department of Forensic Chemistry, National Sun Yat-sen University, Kaohsiung, Taiwan; ²Institute of Forensic Chemistry, National Sun Yat-sen University, Kaohsiung, Taiwan

Keywords: correlation coefficients, analytical methods, forensic toxicology, forensic chemistry

In the present study, we evaluated a public correlation between serum and plasma levels of matrix metalloproteinase 9 (MMP-9) in forensic toxicology. The authors reported that the correlation coefficient between serum and plasma levels of MMP-9 is 0.90. However, the authors did not mention the sampling conditions of the samples. The authors also stated that the samples were collected from the same individuals under different sampling conditions. This is a methodological weakness in using correlation coefficients to assess the interchangeability of analyte data between samples collected under different sampling conditions. The authors should have mentioned the sampling conditions of the samples. The authors should also have mentioned the sampling conditions of the samples. The authors should also have mentioned the sampling conditions of the samples. The authors should also have mentioned the sampling conditions of the samples.

Private Queries?



Does Joe
have
cancer?

query

answer



Letter to the Editor

Methodological weakness in using correlation coefficients for assessing the interchangeability of analyte data between samples collected under different sampling conditions - the example of matrix metalloproteinase 9 determined in serum and plasma samples

Wen Jiang^{1,2} and Chao-Yang Wu^{3*}

¹Department of Clinical Chemistry, University of Medicine and Health Sciences, Huzhou, China; ²Department of Clinical Chemistry, Huzhou University, Huzhou, China; ³Department of Clinical Chemistry, Huzhou University, Huzhou, China

Keywords: correlation coefficients; matrix metalloproteinase 9; serum; plasma

In this paper, Jiang et al. (1) recently reported a method for assessing the interchangeability of analyte data between samples collected under different sampling conditions. The authors used correlation coefficients to assess the interchangeability of analyte data between samples collected under different sampling conditions. However, the use of correlation coefficients to assess the interchangeability of analyte data between samples collected under different sampling conditions is methodologically weak. The authors used correlation coefficients to assess the interchangeability of analyte data between samples collected under different sampling conditions. However, the use of correlation coefficients to assess the interchangeability of analyte data between samples collected under different sampling conditions is methodologically weak. The authors used correlation coefficients to assess the interchangeability of analyte data between samples collected under different sampling conditions. However, the use of correlation coefficients to assess the interchangeability of analyte data between samples collected under different sampling conditions is methodologically weak.

Private Queries?



Does Joe
have
cancer?



The Journal of Clinical Chemistry, Volume 50, Number 10, October 2010, pp 1611-1612

Letter to the Editor

Methodological weakness in using correlation coefficients for assessing the interchangeability of analyte data between samples collected under different sampling conditions – the example of matrix metalloproteinase 9 determined in serum and plasma samples

Wen Jiang^{1,2} and Chao-Yang Ho^{1,2}

¹Department of Clinical Chemistry, ChangGung Memorial Hospital, Taipei, Taiwan
²Institute of Biotechnology, National Tsing-Tung University, Hsinchu, Taiwan

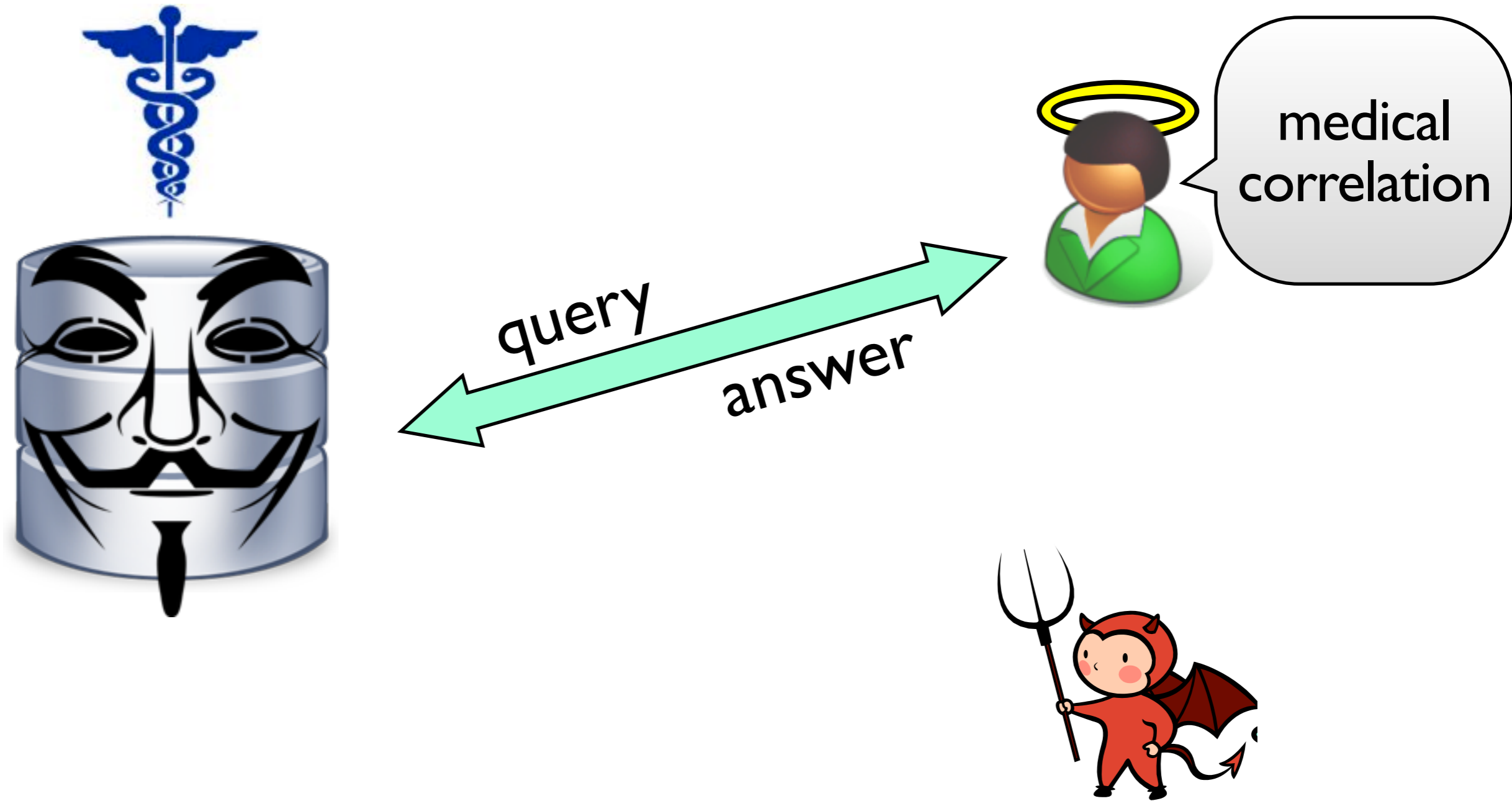
Keywords: correlation coefficients, interchangeability, methodological weakness, sampling conditions

In this journal, Jiang et al. (1) recently reported a method for assessing the interchangeability of analyte data between samples collected under different sampling conditions. The authors used correlation coefficients to assess the interchangeability of analyte data between samples collected under different sampling conditions. However, the use of correlation coefficients to assess the interchangeability of analyte data between samples collected under different sampling conditions is methodologically weak. The authors used correlation coefficients to assess the interchangeability of analyte data between samples collected under different sampling conditions. However, the use of correlation coefficients to assess the interchangeability of analyte data between samples collected under different sampling conditions is methodologically weak. The authors used correlation coefficients to assess the interchangeability of analyte data between samples collected under different sampling conditions. However, the use of correlation coefficients to assess the interchangeability of analyte data between samples collected under different sampling conditions is methodologically weak.

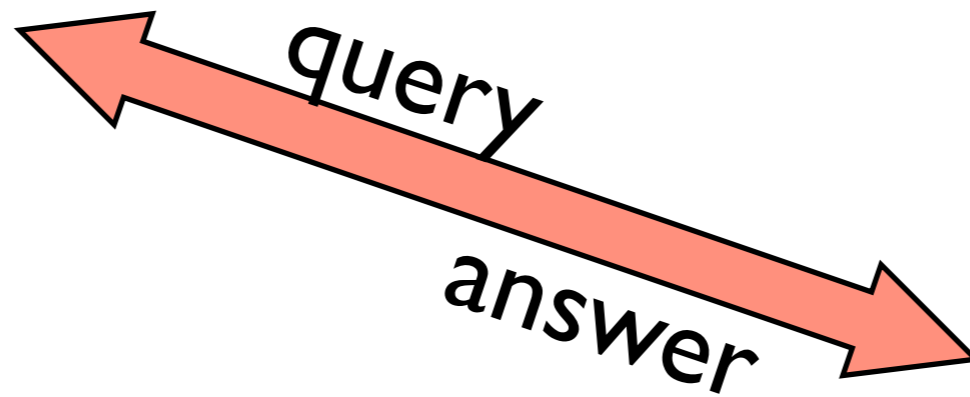
Anonymization?



Anonymization?



Anonymization?



Anonymization



Anonymous Data



Anonymization



Anonymous Data



Anonymization



Additional Data



Anonymous Data



Anonymization



Additional Data



Anonymous Data

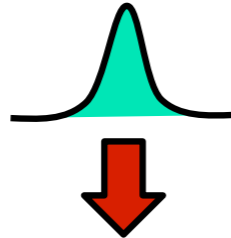


A Possible Solution: Differential Privacy

Differential Privacy



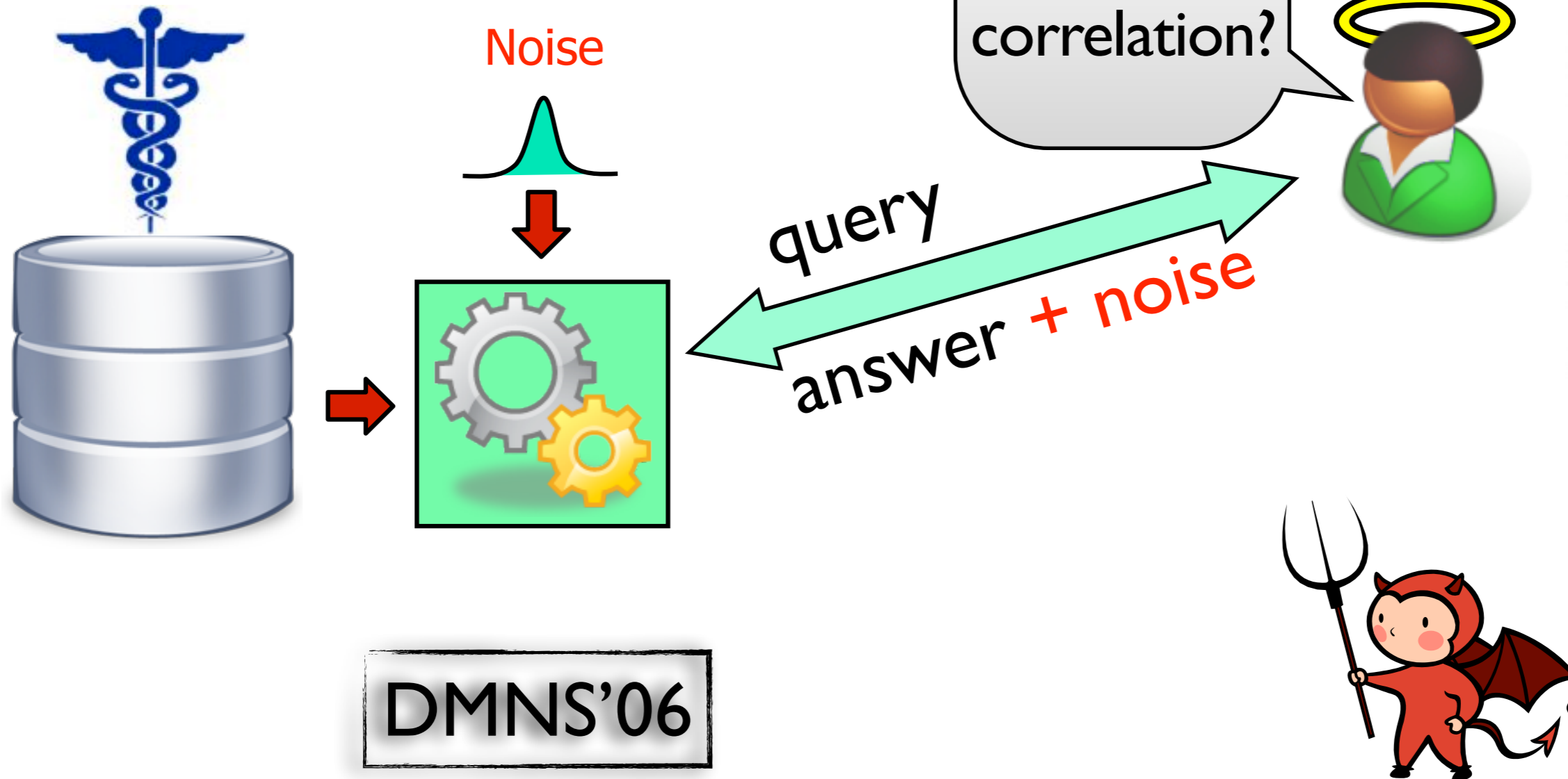
Noise



DMNS'06



Differential Privacy



Letter to the Editor

Methodological weakness in using correlation coefficients for assessing the interchangeability of analytic data between samples collected under different sampling conditions - the example of matrix metalloproteinase 9 determined in serum and plasma samples

Alan Jang^{1,2*} and Chao-Yang Ho^{2*}

¹Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA; ²Department of Biostatistics, Harvard School of Public Health, Boston, MA, USA

*Correspondence: Alan Jang, Department of Biostatistics, Harvard School of Public Health, 665 Huntington Avenue, Boston, MA 02115, USA; Email: alanj@hsph.harvard.edu

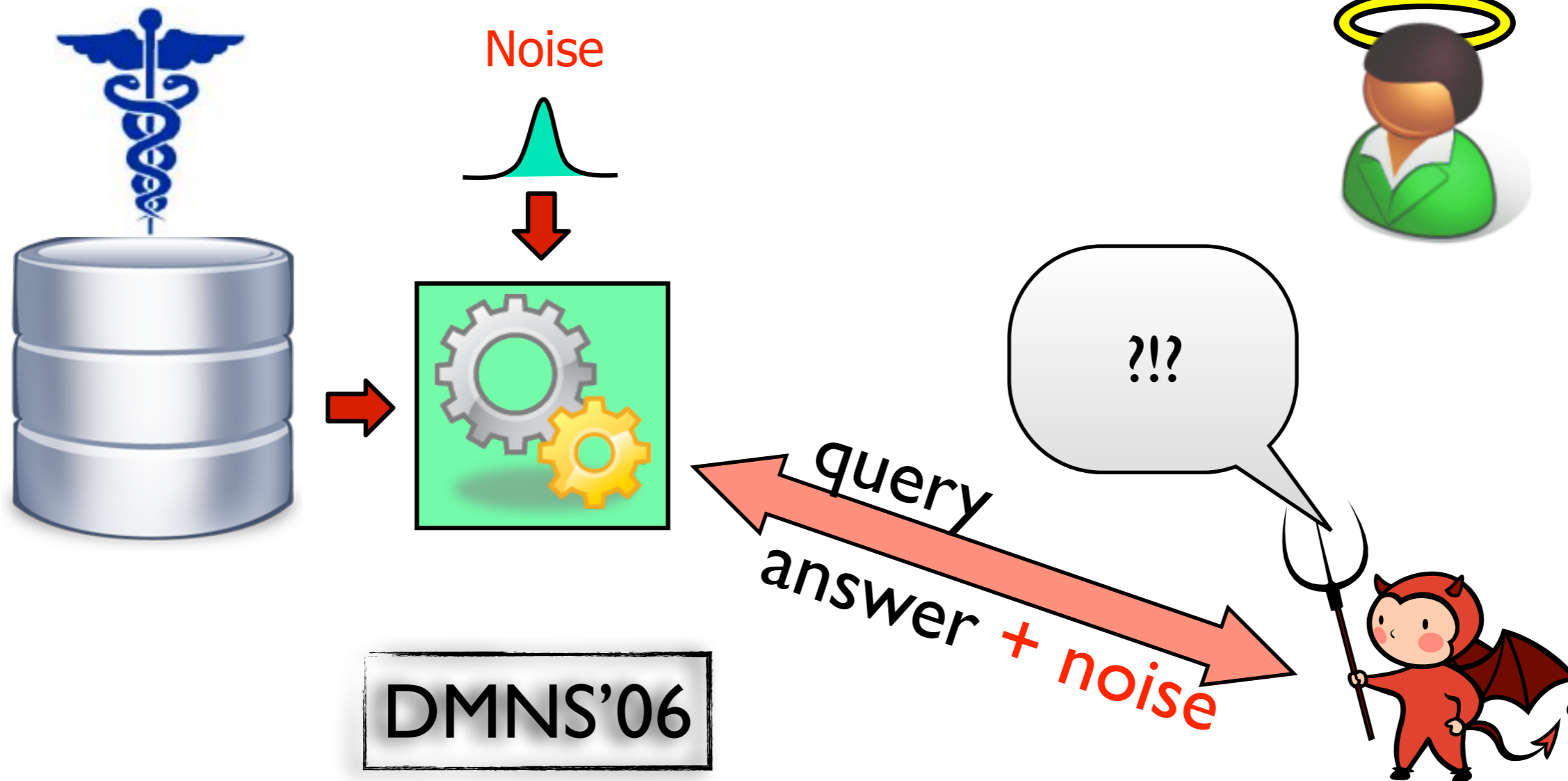
Received: 15 October 2014; Accepted: 20 November 2014; Published: 24 November 2014

Cite this article as: Jang A, Ho C: Methodological weakness in using correlation coefficients for assessing the interchangeability of analytic data between samples collected under different sampling conditions - the example of matrix metalloproteinase 9 determined in serum and plasma samples. *BMC Medical Research Methodology* 2014 **14**:111

Keywords: differential privacy, matrix metalloproteinase-9, correlation coefficient

In this paper, we discuss the methodological weakness in using correlation coefficients to assess the interchangeability of analytic data between samples collected under different sampling conditions. We illustrate this weakness using the example of matrix metalloproteinase 9 (MMP-9) determined in serum and plasma samples. We show that the use of correlation coefficients to assess the interchangeability of analytic data between samples collected under different sampling conditions is methodologically weak because it does not account for the possibility of differential privacy. We propose a method to assess the interchangeability of analytic data between samples collected under different sampling conditions that accounts for the possibility of differential privacy. We show that this method is methodologically sound and can be used to assess the interchangeability of analytic data between samples collected under different sampling conditions.

Differential Privacy



Letter to the Editor

Methodological weakness in using correlation coefficients for assessing the interchangeability of analytic data between samples collected under different sampling conditions - the example of matrix metalloproteinase 9 determined in serum and plasma samples

Kwan Jung^{1,2} and Chee-Hyung Kim^{1,2}

¹Department of Biostatistics, Seoul National University, Seoul, Korea; ²Seoul National University Medical Research Center, Seoul, Korea

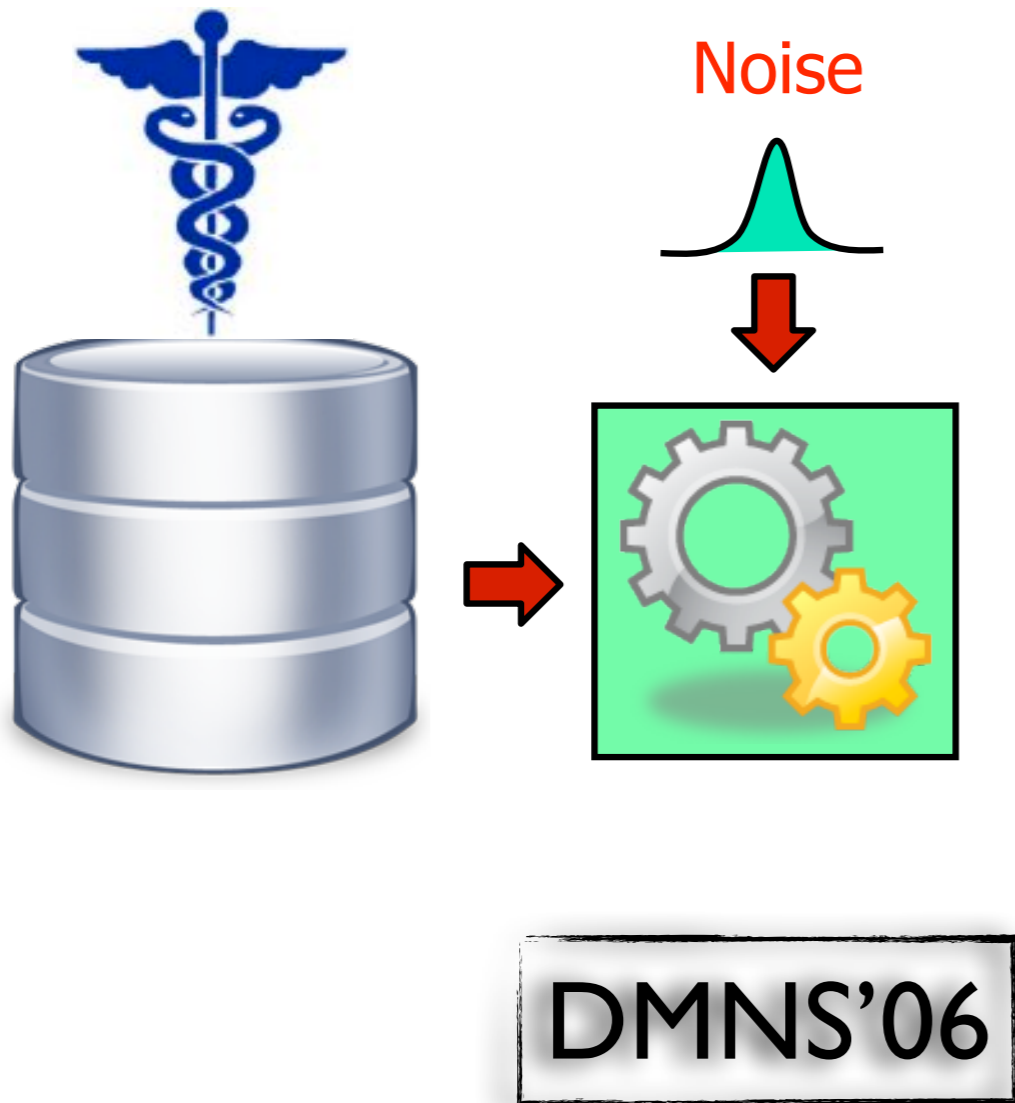
Received 15 September 2006; accepted 20 October 2006

The authors thank the reviewers for their constructive comments and suggestions. This work was supported by the National Natural Science Foundation of China (Grant No. 81273055) and the National Natural Science Foundation of Korea (Grant No. 2003-000-000-10000).

Correspondence: Kwan Jung, Department of Biostatistics, Seoul National University, 51-1 Yonsei-ro, Seoul 151-747, Korea. Tel: +82 2 762 7142; Fax: +82 2 762 7142; Email: kwangjung@plaza.snu.ac.kr

© 2006 Blackwell Publishing Ltd, *Journal of Clinical Pharmacy and Therapeutics*, 31, 541-542

Differential Privacy



Letter to the Editor

Methodological weakness in using correlation coefficients for assessing the interchangeability of analyte data between samples collected under different sampling conditions - the example of matrix metalloproteinase 9 determined in serum and plasma samples

Alan Jung^{1,2*} and Chao-Yang Ho^{2*}

¹Department of Biotechnology, National Sun Yat-sen University, Keelung, Taiwan; ²Department of Biotechnology, National Sun Yat-sen University, Keelung, Taiwan

**Correspondence: Alan Jung, Department of Biotechnology, National Sun Yat-sen University, Keelung, Taiwan. E-mail: alanjung@ms.nsysu.edu.tw*

Received 15 November 2011; accepted 15 November 2011; published online 15 November 2011

Keywords: correlation coefficient; matrix metalloproteinase 9; interchangeability

In the present study, we evaluated the interchangeability of analyte data between samples collected under different sampling conditions. The interchangeability of analyte data between samples collected under different sampling conditions was assessed using correlation coefficients. The results showed that the correlation coefficients were significantly lower than those reported in the literature. This finding suggests that the use of correlation coefficients to assess the interchangeability of analyte data between samples collected under different sampling conditions is not appropriate. We recommend that other methods be used to assess the interchangeability of analyte data between samples collected under different sampling conditions.

Fundamental Law of Information Reconstruction

The release of **too many** overly **accurate** statistics gives privacy violations.



[DinurNissim02]

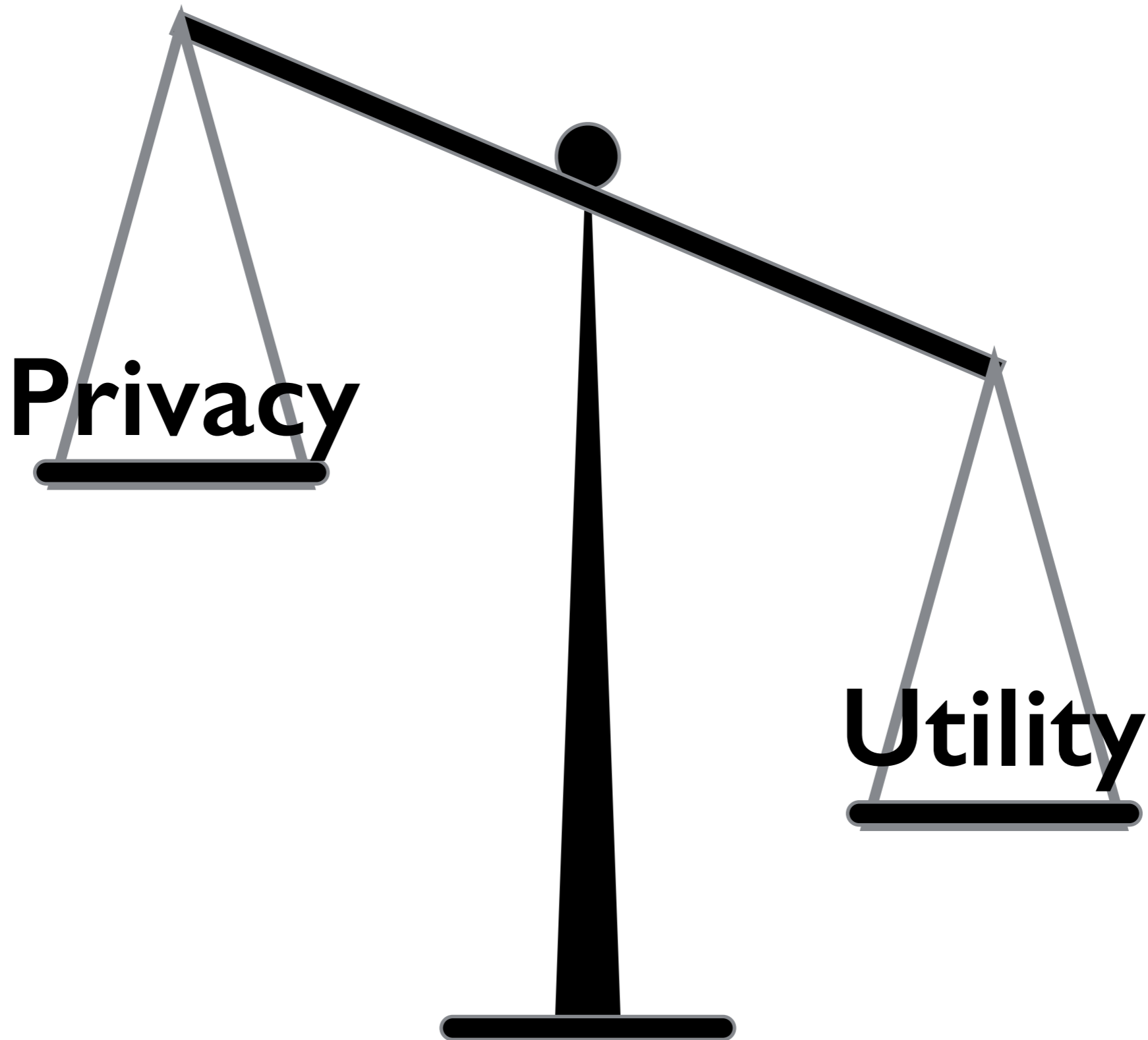
Privacy vs. Utility

Privacy vs. Utility

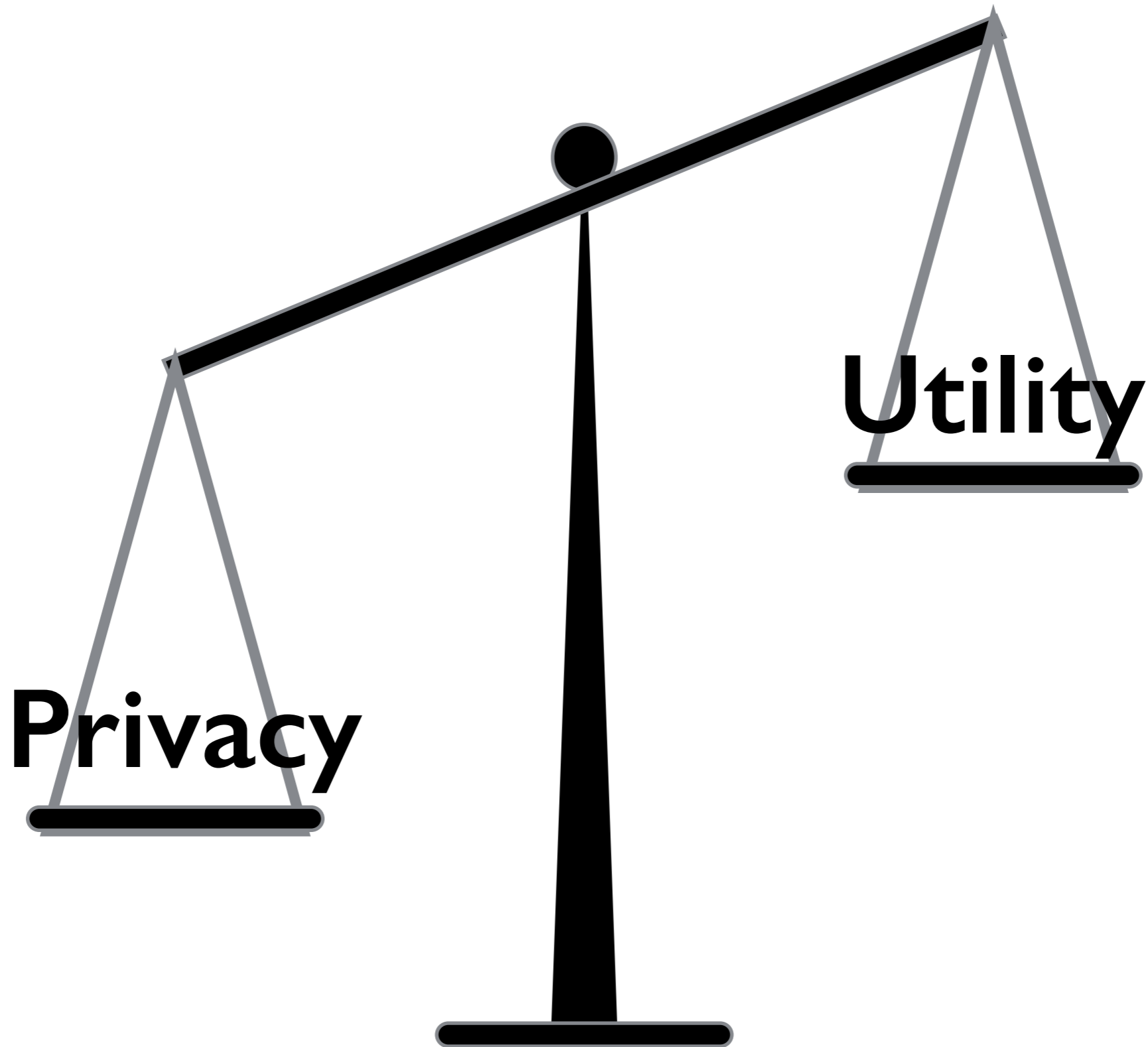
Privacy

Utility

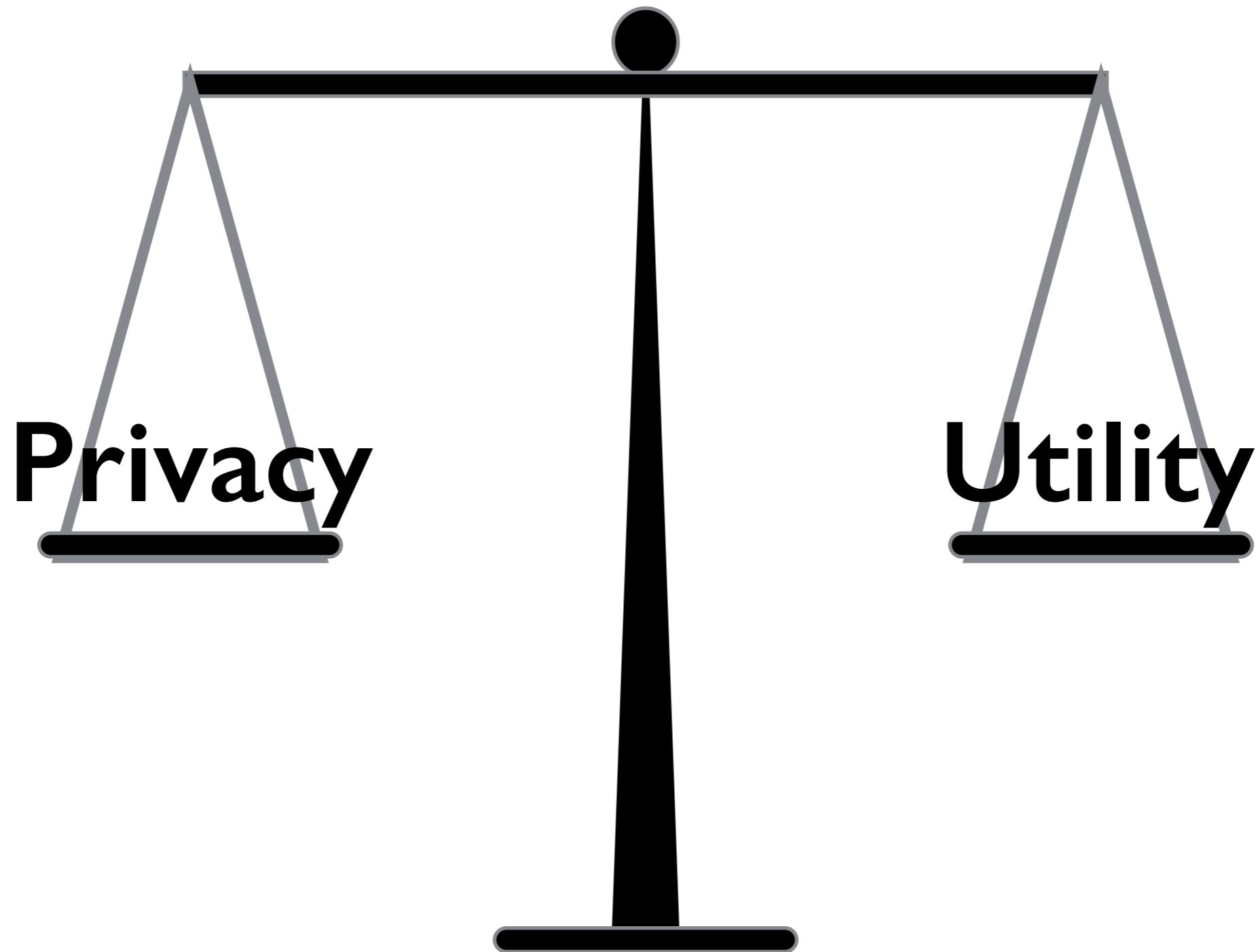
Privacy vs. Utility



Privacy vs. Utility



Privacy vs. Utility



This class:
understanding the mathematical and
computational meaning of this trade-off.

Some Official Users

- US Census Bureau - onTheMap, new releases
- Google - RAPPOR tool for Chrome
- Apple - typing statistics reports
- LeapYear - startup
- ...

Syllabus for the course

Location: Davis 113A

Time: Thursday 16:30 - 19:00?

Office Hours: Friday 11:00 - 12:00 or by appointment

Discussion forums: NB and Piazza

class website: <http://www.buffalo.edu/~gaboardi/teaching/CSE711-spring17.html>

Course load:

- presenting part of the material,
- commenting on the material presented every week, beforehand on NB and during class,
- working on a project and presenting the results (optional)

Grading

50% - material presentation

50% - engagement and participation in class and on NB and Piazza

Schedule

Date	Topic	Presenter
2/02	Course Overview and Introduction to Differential Privacy SV Chapter 1	Marco Gaboardi
2/09	Composition Theorems for Differential Privacy SV Chapter 2	Marco Gaboardi
2/16		
2/23		
3/02		
3/09		
3/16		
3/23	No class - Spring Break	
3/30		
4/06		
4/13		
4/20		
4/27		
5/04		
5/11		

Reference Material

Cynthia Dwork and Aaron Roth,
“The Algorithmic Foundations of Differential Privacy,” 2014
Linked from the class website.

Salil Vadhan,
“The Complexity of Differential Privacy” 2016.
Linked from the class website.

Questions?