ABSTRACT
Comorbidities of obesity were ranked by risk ratio, using the MeSH descriptors associated with articles in the PubMed corpus.

METHODOLOGY SUMMARY
Potential comorbid conditions were gleaned from the MeSH 2015 vocabulary file. Those MeSH descriptors whose semantic types were 'Disease or Syndrome' (T047) or T047's children ('Mental or Behavioral Dysfunction', T048; and 'Neoplastic Process', T191) were extracted using a preprocessor and written to an output file (see MeSHPruner.java and src/main/resources/MeSH_Disease_Or_Syndrome.xml ; note that the output file does not conform to the MeSH 2015 DTD since it was transformed using XSLT into a more manageable format.) This preprocessing identified 3,679 diseases or syndromes.

A risk ratio for the presence of each comorbid condition was calculated. A risk ratio is the ratio of the probability of an event occurring in the presence of an influencing factor to the probability of an event not occurring, given the influence of the same factor. Determining these probabilities requires four counts, which were determined using the following four Entrez queries:

a. Obesity AND comorbid condition (both conditions are present);
b. Obesity NOT comorbid condition (only obesity is present;
c. Comorbid condition NOT obesity (only the comorbid condition is present); and
d. The total number of documents in the PubMed corpus, minus the count from query 'a' (i.e., neither condition is present)

These counts yield the risk ratio via the following expression: (symbols refer to the Entrez queries defined above)

$$RR = a/(a + b) \,/\, c/(c + d)$$

The denominator of this equation is often called the "control group". In our analysis this control group contains all the PubMed articles that did not have the "obesity" descriptor in their major heading; i.e., it contains those articles in which obesity was not a major topic.

DISCUSSION
In the PubMed corpus, those articles that discussed obesity nearly always also discussed nutritional disorders. Of the twelve descriptors with the largest risk ratio, eight discuss nutrition or glucose metabolism and another refers to a congenital disorder that affects metabolism. In other words, writers contributing to the journals indexed by PubMed overwhelmingly believe obesity is associated with metabolism.

Descriptor	Risk Ratio
Overnutrition	264937.09
Nutrition Disorders	1018.03
Binge-Eating Disorder	228.17
Angelman Syndrome	138.14
Nutritional and Metabolic Diseases	98.77
WAGR Syndrome	92.16
Hyperinsulinism	91.67
Metabolic Syndrome X	81.54
Acanthosis Nigricans	80.55
Laurence-Moon Syndrome	71.04
Child Nutrition Disorders	65.84
Glucose Intolerance	56.61

In this analysis there is no way to determine which disease is primary: is disease A comorbid with disease B or the other way around? For a more concrete example, consider 'overnutrition': when this analysis is run using this MeSH descriptor you find type two diabetes mentioned 35 times more often, compared to the control group. But is diabetes comorbid with overnutrition? Or are these articles discussing overnutrition as a comorbid condition of diabetes? Without this information the data is subject to the biases of the person reading it.

Joseph Rose, Nov. 2015