

# Intuition versus Mathematics: The Case of the Hospital Problem

Jane Watson

*University of Tasmania*  
<Jane.Watson@utas.edu.au>

This paper considers 33 preservice secondary mathematics teachers' solutions to a famous sampling problem, well known for confounding educated adults. Of particular interest is the Bachelor of Teaching students' use of intuition and/or formal mathematics in reaching a conclusion. The relationships of solution strategy to students' background in formal mathematics and to gender are also considered. Implications for teaching statistics at both the secondary and preservice teacher education levels are discussed briefly.

A town has two hospitals. On the average, there are 45 babies delivered each day in the larger hospital. The smaller hospital has about 15 births each day. Fifty percent of all babies born in the town are boys. In one year each hospital recorded those days in which the number of boys born was 60% or more of the total deliveries for that day in that hospital.

Do you think that it's more likely that the larger hospital recorded more such days, that the smaller hospital did, or that the two recorded roughly the same number of such days?

The *hospital* problem, as the above problem is usually named, began its life in the work of Kahneman and Tversky (1972), in their ground-breaking work on people's understanding of representativeness. Although there have been criticisms of the wording used by Kahneman and Tversky (e.g., Well, Pollatsek, & Boyce, 1990), this problem has been used repeatedly over the years to explore understanding of sampling and the variation associated with different sample sizes (cf. Well et al., 1990). Tversky and Kahneman (1971) found that university students were susceptible to the "law of small numbers", that is, the belief that a small sample represents a population just as well as a large one. For the problem above this leads to the conclusion that the two hospitals recorded about the same number of days with 60% or more boys born. Although school students have not often been asked this question, Fischbein and Schnarch (1997) found only one student of 60 in grades 5, 7, and 9 who suggested that the smaller hospital would have the more extreme result. Watson, Collis, and Moritz (1995), in a study of students' understanding of sampling in grades 3, 5, 7, and 9, found that of 12 students asked a similar question, none could argue from a correct basis. Watson and Moritz (2000) asked 59 students the same problem and of these, only six recognised the smaller sample as more likely to have the extreme result and gave adequate justification. With younger children the choice of the large or small sample appeared random, the reasons given for choosing the small sample could also be given for choosing the large one, and story-telling often dominated answers. Reasons given for saying the large and small samples had the same chance of having the extreme result included "because it's random" or "because percentages aren't definite numbers and so are the same for each sample".

Researchers studying people's appreciation of sample size in the context of the hospital problem, however, have not explored the understanding of people with a formal background in mathematics and the potential to call upon formal statistics to answer the hospital question in terms of the actual probabilities associated with the two samples having 60% or more boys. Anecdotal evidence (N. Windsor, personal communication, November 9, 1998) indicates that many senior secondary students, when given the hospital problem immediately after a unit on

the normal approximation to the binomial distribution, do not recognise its applicability. Of those who do, many still have difficulty carrying out the process correctly to achieve the appropriate conclusion.

Whereas perhaps not all preservice mathematics teachers would have been exposed to the theoretical mathematics necessary to calculate the probabilities in each hospital, at least it would be expected that they would have access to the basic mathematics to calculate percentages, fractions, and basic probabilities based on equal likelihood (of births of boys and girls). For these people, it is hence of interest to observe what general mathematical knowledge is used to support an intuitive suggestion. It is also of interest to see which mathematical skills are chosen by those with more formal training and whether this is backed up with any reference to intuition. Some people may have experienced sampling in their personal or professional lives that helps them appreciate the likelihood of greater variation in the smaller sample and feel no need for further justification except for a statement. Others may have the ability to calculate a normal approximation to the binomial distribution without any appreciation for the meaning of the result.

The opportunity to explore further the use of formal mathematics as a solution strategy for the hospital problem arose in the context of a case study used in a preservice program for secondary mathematics teachers. The case study, *Chances Are*, was one of a series published by Harvard University to be used with high school mathematics teachers (Merseeth & Karp, 1997). The pre-case worksheet for the case study presented the problem as given at the beginning of this paper and asked participants to provide a solution and list any assumptions made. This case study was used with three groups of Bachelor of Teaching (BTeach) students.

The objective in this small study is to add to the information available on strategies adults, in particular people who are planning to be high school mathematics teachers, will choose to answer the hospital problem. The aims are (a) to observe whether mathematics, intuition, or a combination of the two is most successful in achieving an acceptable solution, (b) to document the types of mathematics employed to support intuition or as the basis for a mathematical justification, (c) to suggest propensities for solution strategies of students with more or less formal mathematics training or by gender, (d) to report students' reactions to the problem, and (e) to suggest action to improve both intuitions and confirmatory skills in this area.

## Method

The 33 preservice secondary mathematics teachers in this study were aged from 21 years, having just finished a university degree, to middle age, having worked in other careers before deciding to become mathematics teachers. The older students came from careers in geology, biology, engineering, meteorology, statistics, and taxi driving. The group varied widely in its previous exposure to formal mathematics courses. One was on leave from enrolment in a PhD in mathematics, whereas others had done no formal mathematics at university but had convinced the head of school (not the author) that they had a mathematics background adequate for the course. Twenty-three BTeach students were classified as having at least a minor in mathematics and 10 were classified as "other". There were 14 females and 19 males in the sample. Although this was not a random sample of preservice secondary mathematics teachers across Australia, it is thought to be representative of the types of students in preservice mathematics courses.

The BTeach students were given the pre-case worksheet to complete overnight before the case study was handed out. They were told to answer the question by themselves and that, although the solutions would be collected and read by the lecturer, no marks would be awarded, except for completing the exercise. The students knew that they would be participating in a case study based on the problem and appeared to be motivated to think seriously about it. Only one solution was very short and appeared to be by a student who had taken the problem less seriously or had little idea of how to solve it.

The method of analysis employed was a clustering technique (Miles & Huberman, 1994), which looked at similarities in solutions to the problem. Of particular interest were those judged to be based on intuition in relation to sample size with no support in the way of mathematical calculations, those that displayed mathematical calculations but displayed no indication of intuition influencing the answer, and those that appeared to mix intuition with mathematical calculations. Further, when mathematics was employed in the solution, it was of interest whether it might be termed basic mathematics, such as percentages and fractions, or theoretical mathematics based on the binomial and/or normal distributions. Finally it was of interest whether the conclusion reached was correct or incorrect. The analysis was based on “observed” outcomes, that is, only what was written on paper. If a student held an intuition that was not stated, obviously no credit was given. After the clustering analysis had taken place, note was made of gender and whether students were considered to have a strong background in mathematics.

## Results

The results are summarised in Table 1 by whether the solution to the hospital problem was based on intuition as stated by the student, whether the presentation was based solely on a mathematical argument, or whether a mixture of approaches was stated. The results are also split by the correctness of the conclusion stated, except in one case. In that case a student stated a proposition of equal outcomes, attempted to calculate probabilities, made an error that indicated the small hospital had a very slightly higher probability, and chose that hospital. Given the numerical result the student should have retained the decision of “roughly equal” and was allocated as if that were the conclusion. Examples that illustrate the categories will be given according to the method of approach.

Table 1  
*Clusters of Responses to the Hospital Problem*

Conclusion	Strategy		
	Intuition	Mathematics	Intuition and Mathematics
Correct (n = 18)	2 males	<i>Formal maths</i>	<i>Formal maths</i>
	5 females	1 male, 1 female <i>Basic maths</i> 6 males	2 males, 1 female
Incorrect (n = 15)	4 males	<i>Formal maths</i>	<i>Basic maths</i>
	4 females	3 males, 1 female	1 male, 2 females
	Total = 15	Total = 12	Total = 6

### *Intuition Only*

Fifteen BTeach students gave responses deemed to be based on intuition. Although some of these solutions contained numbers, these were not used with calculations to support the conclusion. Of these 15 solutions, 7 chose the smaller hospital based on the correct reasoning of sample size. These responses appeared to reflect previous experience of some sort involving variation in samples. A typical response is the following, which although it included numbers had no calculations.

I can't work this out mathematically, but intuitively I would say that the smaller hospital is more likely to have more variation from the average of 50% boys. This is because even though the probability of a head (or tail) when tossing a coin is  $1/2$ , this is only the case the more times we toss the coin (ie the head won't necessarily turn up once in every 2 tosses, but the more times the coin is tossed, the closer the heads get to a frequency of a  $1/2$ ). Similarly, the more babies there are, the more likely the frequency of boys is to converge to 50% of births.

One person in this group mentioned that data would approximate a normal distribution but expanded only to say, "the degree of fit to the normal curve can be expected to be better for a larger sample size". The statement appears to reflect previous experience but is not backed up with any further mathematical argument.

Of the eight solutions that did not choose the small hospital, two did not reach a discernible conclusion, either stating two possibilities (larger hospital closer to the norm but both hospitals roughly the same over a year) or only eliminating one (larger would not have more). One student stated the following as a final comment: "If more births occur at the larger hospital, there is a greater likelihood that more boys will be delivered at the larger hospital." Although this statement is technically correct and would not preclude a correct comment on the hospitals, no other comment was made. The other five students indicated that the hospitals would have roughly the same number of days with 60% or more boys born. Typical of these is the following.

Percentages are proportions of original sizes. 60% is the same for both hospitals – only depends on original number to work it out. There is no reason why the larger hospital should have 60% boys more often than the smaller hospital. It's not asking which hospital turns out more boys because then the bigger hospital would do that.

Of the 15 BTeach students who used intuition only as a basis for their solutions, 8 were among the group with less theoretical mathematics in their undergraduate degrees. Of these 8, half reached the correct conclusion. Of the 15, 6 were male and 9 were female with the females being more successful (56%) than the males (33%).

### *Mathematics Only*

Twelve students provided solutions based on mathematical arguments with no indication of previous experience or intuition that would lead to belief in *that* solution at the start of the exercise. Eight of these twelve students reached the correct conclusion. Two of these students recalled from previous experience that the problem involved the binomial distribution and went to their old texts to find out how to carry out the calculations. One used the normal approximation to the binomial for both hospitals, whereas the other used binomial tables for the small hospital and the normal approximation for the large one. Neither person commented on their surprise or otherwise with the answer, although the person providing the second response indicated an initial belief that the hospitals would have roughly the same number of days.

One other student attempted to calculate binomial probabilities but only calculated the exact probabilities for 9 boys out of 15 and 27 boys out of 45. Although incorrect to answer the question based on “60% or more boys”, these values led to the correct conclusion that the smaller hospital was more likely to have exactly “60% boys” born.

The other four students with the correct conclusion used simpler mathematical arguments based on how likely it would be to achieve deviations from the expected number of births to reach the 60% value. After stating that for the total of 60 births per day, 30 on average would be boys, one student argued as follows.

Given that, for the larger hospital to record 60% or greater boys in a given day would require at least (27/30) 90% of the average number of boys born per day, as compared to only 30% (9/30) or more for the smaller hospital, one would suggest that it is more likely for the smaller hospital to record a greater number of such days.

After displaying a few percentage calculations, another student argued as follows.

Given that 53.3% of the boys are already present (8 for the small hospital and 24 for the large hospital). It only requires 1 more boy for the small hospital to give 60% but requires 3 boys to give 60% for the large hospital. The chances of 1 more boy is [sic] higher than the chances of 3 more boys.

Variations on this theme were expressed with probabilities, the percentages represented by each birth, and the days required by the large hospital to redress an imbalance.

Four of the students who used formal mathematics in their solutions made errors that led them to the wrong conclusion. One made an error in the formula when applying the normal approximation to the binomial and one applied a hypothesis test incorrectly to the two hospitals. Both concluded that the hospitals would be equally likely to have 60% or more boys born. A third student did not think there was enough information available to use the binomial distribution, did some calculations to estimate total births with “unknown” probabilities for “number of days greater than 60%”. This led to the conclusion that the large hospital would have more “60% or greater” days. The final solution in this group was one mentioned earlier where an incorrect calculation of probabilities (0.21 and 0.23) should have led to the conclusion that the hospitals would have roughly the same number of days with 60% or more boys born.

Only one of the students in this group of 12 was considered to have a weak formal mathematics background. This was the student who used the normal approximation correctly for both hospitals. His university major was in zoology and although he had done no formal mathematics courses at university, he had worked for 10 years in positions involving scientific research. Of this group of 12 using only mathematics for the problem, two were female, one providing a correct solution. Of the 10 males, 7 were correct.

### *Intuition and Mathematics*

Only six students provided solutions that appeared to attempt to justify intuition or previous experience with a mathematical argument, of which half were correct. The three BTeach students providing correct conclusions used very different approaches. One provided a graph of a normal distribution curve centred at 50, with the horizontal axis labelled “percentages of boys” and the vertical axis labelled “Num. of days”. Under the graph was a paragraph, edited extracts of which are as follows.

If all things are considered equal both should produce the above distribution. However with the smaller hospital I would imagine there is more chance of (looking at extremes) all babies being boys than in the other hospital with an average of 45. Hence I would envisage that the curve for the smaller hospital would be a bit flatter resulting in ... a lesser value for the 60% mark ... so although it would be flatter it would be longer hence we may find that the larger hospital with its larger sample will result in more

days closer to 50% resulting in a more pointy distribution while the other is more flat - at this point I would imagine that the smaller hospital may lean towards a greater number.

The second student reported an a priori belief in equality of the hospitals, but then argued that as  $n$  approached infinity the ratio of boys to all births would approach  $1/2$ , hence “the smaller hospital will be more likely to have 60% or more boys a day over a year”. To confirm this belief the student went to a senior secondary mathematics book, found the formula for the binomial distribution and correctly calculated the probabilities for both hospitals using a calculator! The third student expressed the correct intuition based on sample size and attempted to follow the approach of the previous student. Without a reference for the appropriate formula, the student guessed and achieved the correct relationship of the two probabilities but realised that the actual values were incorrect.

The other three students who mixed intuition and mathematics used basic mathematics. All concluded that the hospitals would have about the same number of days with 60% or more boys born. One argued sequentially and erroneously, that no matter how many children are born, the probability of all of them being boys is 50%. The claim was made individually for up to four babies and then extrapolated to 15, 45, and 60. Another worked with fractions, concluding the number of ways to get 60% or more boys was  $6/15 = 2/5$  in the smaller hospital and  $18/45 = 2/5$  in the larger hospital, and hence “the two hospitals record roughly the same number of days”. His final comments, however, indicate that he may have become confused by the added information about the 365 days on which data were recorded. He stated the following.

Initially I believed that the smaller hospital has a greater likelihood of having  $\geq 60\%$  of births being boys because of the greater variation due to the small sample size.

However because the births are independent and at each one the  $P(B)=1/2$ , I believe that over the period of 365 days (which is a fair amount of trials) the variation will average out so that the two hospitals will record roughly equal number of days where the percentage of boys born was  $\geq 60\%$ .

Finally, after translating percentages into numbers for the two hospitals and doing comparisons, a student argued, “For the small hospital the fluctuation between 50-60% is only 1-2 children, and 4-5 for the large hospital but they [large] have more children so I think it would be about the same for both.”

Only one of the six students in this group was considered to have a weak formal mathematics background. This was the one who made the erroneous claims about all boys occurring 50% of the time independent of sample size. In this group half were male.

## Discussion and Student Reaction

Several observations can be made about the strategies and success of preservice mathematics teachers in solving the hospital problem. Their success rate (55%) is not surprising considering other results for tertiary students. This did not differ greatly for females (50%) or males (58%). What may be considered surprising is the apparent lack of attempt to engage both intuition (or previous experience) and mathematical skills. Only six students (18%) appeared to be working with both. Those who used intuition only may not have encountered the binomial or normal distributions before but all appeared to understand subsequent discussions of these when presented by other members of the class. They also may not have had immediate access to sources that could supply needed information or felt that the problem was not intended to be *that complicated*. That those who used intuition only were only successful about half of the time points to the need to make students aware of the pitfalls of not looking for mathematics to back up intuitions.

The students who presented mathematical arguments with no hint of using intuition or previous experience were more successful (67%) but those who made errors that led to the wrong conclusion illustrate the difficulty when there is no intuition or previous experience to test against one's mathematical calculations.

Of the six who employed both intuition and mathematics, half reached the correct conclusion. These used formal mathematics related to the model appropriate for the hospital problem. The others based arguments on basic mathematics that was misinterpreted in the complex context. The response that indicated the chances of all-boy samples were 50%, was similar to responses about four coin tosses being tails observed by Moritz and Watson (2000), for between 28% and 49% of students in grades 6 to 11. That a preservice teacher has the same difficulty as many students points to this as an area for research based on instructional intervention.

Of the 10 students considered to have weak formal mathematics backgrounds, eight used intuition or previous experience alone to solve the problem, one used formal mathematics only, and one appeared to mix intuition and basic mathematics (unsuccessfully). Overall half of these students were successful, a result not very different from those considered to have a strong mathematics background (57%).

Student reactions to the correct response during class discussion were mixed with some who had used intuition at first seemingly overwhelmed by the theoretical solution. Several mathematics majors lamented their lack of practical experience in statistics that would have assisted in making judgments about sample size. This was summed up by the PhD student who said the following in his evaluation of the case study.

...there is no substitute for exposure to practical problems when it comes to understanding statistics. Even though I have completed a few courses in statistics at university level I still found that my intuition let me down when it came to working on this problem.

He further claimed in class discussion that he had never carried out any sampling during his study of statistics.

Several students made comments suggesting they had difficulty because the problem stated the extreme result in terms of boys' births only. One claimed that the problem should be reworded to include girls but then went on to solve it correctly using intuition only. Not appreciating the symmetry of the problem in terms of boys and girls was one of the features included in the Harvard case study (Merseeth & Karp, 1997) as a difficulty for high school students. This point was well taken by the preservice teachers and their own difficulties enhanced the relevance of the case study. The fact that a few of the BTeach students experienced the same difficulties with the relationship of percentages and actual numbers as the students in the case study, reinforced the need to concentrate on percentages in the early years of secondary school.

During the class discussion one of the students who had reached the wrong conclusion would not consider other students' correct intuitions until his original mathematics was proved wrong. Having stated " $Bi(n, \theta) \approx N(n\theta, n\theta(1 - \theta))$ ", he forgot the square root for the standard deviation. He thus concluded that  $z = 0.2$  for all  $n$  and hence both hospitals have roughly the same number of days. Once convinced of his error, however, he quickly produced the following argument for the class.

As  $Bi(n, \theta) \approx N(n\theta, \sqrt{n\theta(1 - \theta)})$ , for a value of  $\theta = 0.5$  and an  $x$ -value of .6 or more boys,  $z = (0.6n - 0.5n)/\sqrt{(n(0.5)(0.5))} = 0.2\sqrt{n}$ . Hence as  $n$  increases the  $z$ -value increases. Since we are interested in  $P(z \geq 0.2\sqrt{n})$ , when  $n = 15$ , the probability is greater than when  $n = 45$ .

A suggestion made in the Harvard case study (Merseth & Karp, 1997) was to use 15 coins and 45 coins to simulate the births in the two hospitals. The preservice teachers immediately saw the value of this analogy but also realised that it might be difficult to implement in a classroom. Less noisy ways to provide simulations involved using computer software or a graphics calculator. The probability simulation software for Macintosh, *ProbSim* (Konold & Miller, 1993), provided the opportunity to see samples being created. The outcomes were plotted by the class to document the relative occurrence of extreme results for different sized hospitals to observe the trend with increasing  $n$ . Windsor (1998) provides the instructions necessary to carry out the simulation on a TI-83 and discusses using it with senior secondary students. BTeach students were given the opportunity to trial this method as well. One of the students who had used theoretical mathematics incorrectly produced a spreadsheet simulation in *Excel* that for one year found extreme results occurring for a 15-baby hospital on 118 days, a 30-baby hospital on 68 days, and a 45-baby hospital on 36 days. The PhD student noted above wrote a *Mathematica* script to simulate a year's daily births at each hospital. Running this program for 1000 years produced average values of 42.7 days for the large hospital and 110.6 days for the small hospital for the extreme results. In the end, based on their own experiences and the case study, all students agreed to the importance of using some kind of simulation in conjunction with the theoretical mathematics for the hospital problem.

For people preparing to be mathematics teachers it is disappointing that so few naturally mixed intuition with an attempted mathematical justification. The case study setting of a school class provided an excellent opportunity to discuss this point, as well as the mathematics teachers' role in modelling the relationship between intuition and justification. In the area of chance and data in particular it is essential that preservice teachers be exposed to situations where natural intuition without previous experience may lead to erroneous conclusions. This should reinforce the importance of both types of experiences: intuitive and mathematical. It is only when one has both that one can be sure that the mathematics makes sense.

## References

- Fischbein, E., & Schnarch, D. (1997). The evolution with age of probabilistic, intuitively based misconceptions. *Journal for Research in Mathematics Education*, 28, 96-105.
- Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgement of representativeness. *Cognitive Psychology*, 3, 430-454.
- Konold, C. & Miller, C. (1993). *ProbSim*. [Computer software] Amherst, MA: University of Massachusetts.
- Merseth, K. K., & Karp, J. B. (1997). *Harvard mathematics Case Development Project 1994-1997: Cases of secondary mathematics classrooms*. Cambridge, MA: Harvard Project on Schooling and Children.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Moritz, J., & Watson, J. (2000). Reasoning and expressing probability in students' judgements of coin tossing. *This volume*.
- Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76(2), 105-110.
- Watson, J. M., & Moritz, J. B. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education*, 31, 44-70.
- Watson, J. M., Collis, K. F., & Moritz, J. B. (1995, November). *The development of concepts associated with sampling in grades 3, 5, 7 and 9*. Paper presented at the Annual Conference of the Australian Association for Research in Education, Hobart.
- Well, A. D., Pollatsek, A., & Boyce, S. J. (1990). Understanding the effects of sample size on the variability of the mean. *Organizational Behaviour and Human Decision Processes*, 47, 289-312.
- Windsor, N. J. (1998). Simulating a sampling problem using a TI-83 graphics calculator. *Australian Senior Mathematics Journal*, 12(2), 32-37.