132672

132672

PLANNING OPTIMUM COURT CAPACITY

Bronwyn Lind
Don Weatherburn
Jeanette Packer

NSW BUREAU OF CRIME STATISTICS AND RESEARCH

1990

Published by the NSW Bureau of Crime Statistics and Research
Attorney General's Department
Level 5
20 Bridge Street
Sydney

ISBN 0 7305 8742 8

PREFACE

The NSW Bureau of Crime Statistics has, over the course of its existence, expended considerable effort conducting research into various aspects of the operation of the NSW criminal justice system. In the past that effort has sometimes been directed toward evaluating the response of the courts to legal reform and sometimes (though less commonly) directed toward assessing the impact of legal reform on the operation of the court system.

Both kinds of project are valuable. In their nature, though, neither kind is capable of informing the original decision to change the law or the operation of the justice system. The present report contains the results of one of several new initiatives being undertaken by the Bureau which are designed to provide policy makers with methods by which to identify a need for change as opposed to helping them evaluate the impact of change once it is made.

This report is designed to help managers make more informed decisions about when and how to increase (or decrease) court capacity. Too much court capacity results in a wastage of scarce public resources. Too little court capacity results in court delay and all the inequities associated with it. The problem for court administrators has always been how to find an objective means by which to steer a sensible middle course between the Scylla of court delay and the Charybdis of excess capacity.

It is possible to steer such a course. The present report shows how. The theory underlying the method will appear highly technical to court administrators whose specialisation is in the law rather than in statistics. It is hoped that this will not act as a deterrent. Though full appreciation of the logic underlying the method requires some statistical sophistication, the method itself is quite straightforward and, as is shown in the report, may easily be programmed on a simple desk-top calculator.

Dr Don Weatherburn
Director

June 1991

CONTENTS

SUMMARY

Court administrators frequently have to make judgements about when and how they should increase court capacity in order to meet an increase in demand for court services. A decision which results in too much capacity relative to demand, results in public resources being tied up where they are not needed. On the other hand a decision which results in too little capacity relative to demand, will result in court delay. Decisions about the optimum amount of court capacity, however, are complicated by the fact that there are often marked variations in the duration of each court case and therefore the amount of court time consumed by a number of cases during the course of a year.

This report provides a solution to the problem. It proposes a method which measures the demand for court time as the sum of the hearing times of all cases to be disposed of. The method relies only on the assumption that case hearing times are independent of each other. (In most situations this assumption will be a reasonable one but it will not hold if, for example, the hearing time for a case is dependent on the number of cases awaiting a hearing.)

The method described for choosing the optimum court capacity takes into account both the risk of demand exceeding capacity and the likely amount of spare capacity.

In addition, the report discusses alternatives to changing capacity by assessing the impact of changes in the factors which influence demand for court services. These factors are the number of cases to be disposed of and the mean and variance of case hearing times.

The report describes in detail the mathematical theory underlying the method for determining an optimum court capacity but application of the method involves the use of only a few simple formulae which can easily be programmed on a calculator. The report explains how this is done.

Use of the method, however, does require the measurement of some quantities not normally found in the statistical output of court information systems. Accurate measurements must be taken of (a) the amount of court capacity actually made available for the hearing and disposition of cases, (b) the mean and standard deviation of hearing durations and (c) the number of cases which must be disposed of in a year. The measurement referred to in (c) must take into account the fact that a large number of cases nominally registered for hearing do not actually transpire in hearings, either because the case is 'no-billed' or the accused changes plea, absconds or dies.

1. INTRODUCTION

This report is intended to describe a general method for planning court capacity in relation to the demand for court services and to give some practical illustrations of its utility. There are three fundamental problems to be addressed in planning court capacity. The first involves the question of how much court capacity is required to deal with a given level of demand for court services. The second involves the question of how one should go about determining the future demand for court services. The third involves the question of how court capacity should be spatially distributed, given the known distribution of demand for court capacity. The present report is directed only to the first of these three problems. It is hoped that supplementary reports will address the second and third problems.

At one level it seems an oversimplification to speak of planning court capacity to meet the demand for court services. There are, after all, a multiplicity of such services even within a particular court jurisdiction. There are judges, court rooms, transcription facilities and personnel, stenographers, monitors, legal aid solicitors, sheriffs, chamber magistrates and so on. The level of demand for these services varies considerably, as does the capacity required to meet that demand. Without wishing to understate the difficulties involved in identifying the optimum configuration of these component services for a particular court, it is clear that in the main they are directed to just one object, namely the provision of court time for the hearing and disposition of cases. At this general level the demand for court services in a particular jurisdiction can be understood as the demand for court time in that jurisdiction. Court capacity can then be understood as the total amount of court time made available.

As an approximation, court capacity for a particular class of case can be determined by adding together the sitting times for courts assigned to hearing that class of case. Thus if 10 trial courts sit for 5 days a week during 40 weeks of a year, the theoretical trial court capacity is (10x5x40=) 2000 days.[1] Notice that on this definition court capacity can be expanded in a variety of ways. The obvious way is to build more trial courts. But court capacity can also be expanded by requiring courts to sit for more hours in a day, more days in a week or more weeks in a year. Likewise, in courts which hear a mixture of trials and other kinds of matter, such as sentence matters or civil litigation matters, trial court capacity can be expanded by increasing the proportion of a court's time allocated to the hearing and disposition of criminal trials. This is an important point and ought not to be overlooked in any consideration of exactly how court capacity should be increased.

How do we measure the demand for court time in a particular jurisdiction? It is obviously indexed to some extent by the number of cases which that jurisdiction must dispose of over a specified period, say a year. It should be noted, however, that the number of cases to be disposed of in a year is often a mixture of cases which have arrived for disposal in that year and cases which arrived, but were not disposed of, in previous years. Where there is no backlog the number of cases to be disposed of in a year will be the number of cases arriving in that year. Where there is a backlog, court administrators may plan court capacity to dispose of more cases than arrive in order to reduce the backlog.

Even if the number of cases to be disposed of does not change from one year to the next, the demand for court time can change considerably. The reason for this is obviously that, because each individual case consumes a variable amount of court time, the disposition of a fixed number of cases will take a variable amount of court time.

This last observation suggests that the total demand for court services in a particular year can best be understood as the sum of all the hearing times of cases which are to be disposed of in that year. Of course, since the duration of each case is variable, this sum must vary. If we know the nature of the variation in case duration times, however, it is possible to determine the nature of the variation in the total amount of court time required to dispose of a given number of cases in a particular year. Information on this variation can then be used to select the appropriate court capacity. This is the approach used in this report. The structure of the report is as follows. Section 2 describes the theory and assumptions underlying the method being proposed for determining the court capacity required to meet a given level of demand. Section 3 extends the theory to examine alternatives to changing court capacity. Section 4 describes the practical steps involved in using the method to determine court capacity. Readers interested simply in seeing an application of the method can pass directly to Section 4. Finally, an appendix provides program listings and instructions for use on an HP 32S programmable calculator.

2. THEORETICAL APPROACH TO PLANNING COURT CAPACITY

To develop a full understanding of the method proposed for planning court capacity we need to express the notions of demand for court service and court capacity a little more formally. For the purpose of exposition the following analysis is constructed around the problem of planning District Criminal Trial Court capacity. It will be obvious, however, that the same general principles apply to planning for court capacity in civil or criminal, summary or indictable jurisdictions.

For convenience we will assume that there is no backlog and that the number of cases to be disposed of is measured by the number of cases arriving.

Let t_i denote the hearing time required to dispose of trial i and n denote the number of trials arriving for disposition in a specified period of time, say one year. We assume t_i is a random variable. Thus the demand for court service, D, is simply the sum of the t_i for each of the trials 1 through n, which arrive for disposition in that year. That is:

$$D = t_1 + t_2 + t_3 + \ldots + t_n = \Sigma t_i$$

Since the t_i vary, obviously D must vary from year to year. The nature of the variation in D, moreover, will be determined by the underlying variation in the t_i. Indeed, provided that the t_i are independently distributed with finite means and finite variances, the central limit theorem holds and it can be shown that, for large n, D is distributed normally with mean equal to the sum of the means of the t_i and variance equal to the sum of their variances.[2]

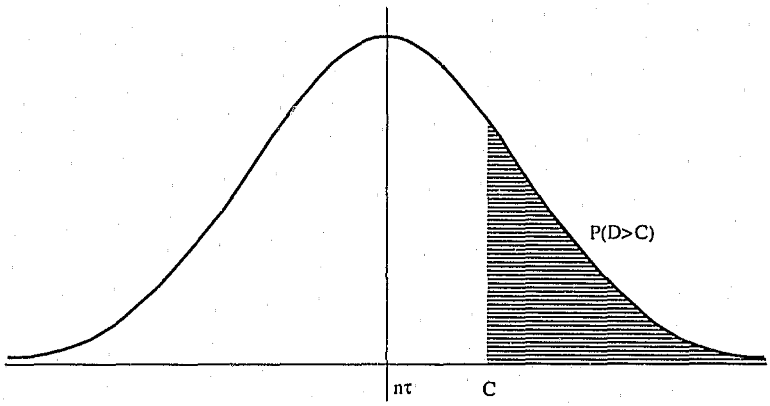
The t_i will certainly have finite means and finite variances as no case will ever be of infinite duration. The assumption that the t_i are independently distributed is also a reasonable one as there is no reason to expect that the length of a trial will depend on the length of any other trial.

Suppose for the moment that the t_i are identically and independently distributed with mean τ and variance σ^2. Then D is normally distributed with mean $n\tau$ and variance $n\sigma^2$.

In other words D will be distributed as illustrated in Figure 1, with frequency distribution:

$$f(D) = \frac{1}{\sqrt{(2\pi n\sigma^2)}} e^{-(D-n\tau)^2 / 2n\sigma^2}$$

Figure 1: Distribution of demand for court time



Court delay will develop whenever a backlog of cases develops. This occurs whenever demand for court time exceeds court capacity. Since D is a random variable, for a given court capacity, C, there will be a probability that demand exceeds capacity, that is, that D>C. It is this probability which is the focus of our attention. We denote it by P(D>C). The shaded area in Figure 1 shows this probability. As can be seen from the figure, the larger the value of C the smaller P(D>C) becomes.

If Z is a standard normal variate (that is, is normally distributed with mean 0 and variance 1) then P(D>C)=P(Z>Q) where

$$Q = (C-n\tau) / \sigma\sqrt{n} \qquad (2.1)$$

For a given probability, r, of demand exceeding capacity, the value of Q such that P(Z>Q)=r can be found in statistical tables of the normal distribution. For example, if r=0.05 then Q=1.645; if r=0.10 then Q=1.282.

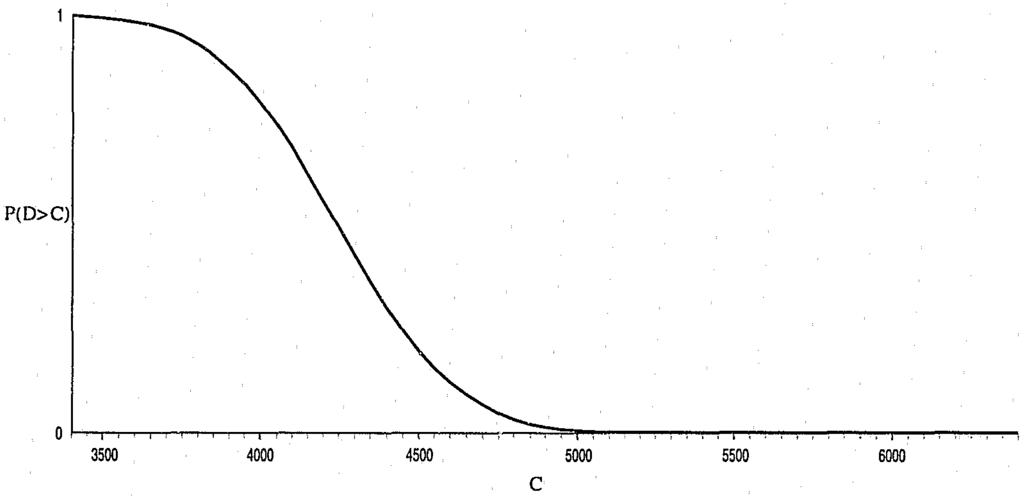
For illustration purposes some example data are used in the remainder of this section of the report. The example data are based on a sample of 190 trial cases drawn from the Sydney District Criminal Court during 1989. The mean and variance of the t_i were estimated from the sample. For our example we shall assume they take the following values (measured in days of court time):

$\tau = 4.2421 \qquad \sigma^2 = 89.9773 = (9.4856)^2$

Let us also assume that there are 1000 cases to be disposed of, that is, that n=1000.

Then D, in our example, is normally distributed with mean 4242 and variance 89977. The functional relationship between P(D>C) and C is shown in Figure 2. For an average demand of 4242 court days, it can be seen from Figure 2 that, if court capacity is less than 3500 court days, demand will almost certainly exceed capacity. At the other extreme there is negligible chance of demand exceeding capacity when capacity is 5000 court days or more.

Figure 2: Probability of demand exceeding capacity, as a function of capacity



CAPACITY REQUIRED

The first problem is to determine what capacity would be required to dispose of a specified number of cases. In theory we can select a capacity so large that there is virtually no chance that demand will exceed capacity. However this approach would not be practical as it would lead to excessive unutilised court time as we shall see. It is therefore necessary to set an acceptable (but not infinitely small) level of risk of demand exceeding capacity.

Suppose we are prepared to accept a risk, r, of demand exceeding capacity. Then, in order to determine what capacity would be required to dispose of n cases for a given risk, r, of demand exceeding capacity, we need to determine the value of C for which P(D>C)=r. We do this by solving expression (2.1) for C, substituting Q_r for Q where Q_r is such that $P(Z>Q_r)=r$. The solution is

$$C = (\sigma\sqrt{n})Q_r + n\tau \qquad (2.2)$$

Using our example data, the capacity required to dispose of 1000 trials with only a 5% risk of demand exceeding capacity is determined as follows. Noting that for r=0.05, Q_r=1.645 and substituting our parameter values in (2.2):

$$C = (9.4856\sqrt{1000}) \times 1.645 + 4242 = 4735$$

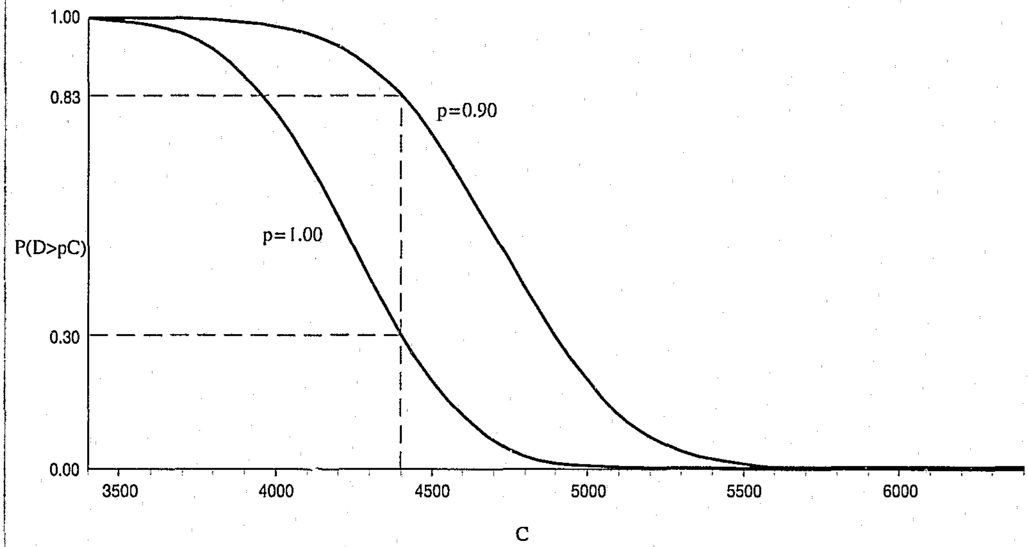
Therefore a court capacity of 4735 court days is required to dispose of 1000 trials with a 5% risk of demand exceeding capacity.

UNUTILISED COURT TIME

Now whenever D<C the courts are underutilised. For efficiency, we wish to maximise the amount of court time utilised while also minimising the risk that D>C. Let pC be the amount of court time used where p is a proportion. Then the complement (1-p)C obviously gives the amount of court time left unused. When demand exceeds pC then unutilised court time is less than (1-p)C. So P(D>pC) gives the probability of unutilised court time being less than (1-p)C. For example, if p=0.90 and demand exceeds 90% of capacity, then less than 10% of capacity is unutilised and the probability of this occurring is P(D>0.90C). Figure 3 plots P(D>pC) against court capacity for p=1.00 and p=0.90. Note that when p=1.00 we have pC=C, so the curve for p=1.00 is the same as that shown in Figure 2.

It can be seen from inspection of Figure 3 that for a court capacity of 4400 days the probability that demand exceeds capacity is 30% and the likelihood that less than 10% of court capacity is left unutilised is 83%. More generally it is evident that, as we reduce the risk that demand for trial court services will exceed capacity and hence reduce the risk of court delay, we necessarily increase the risk of unutilised court time.

Figure 3: Probability of demand exceeding a proportion p of capacity, as a function of capacity



For a specified capacity C we can determine what proportion p of court time will 'almost certainly' be used. Then the amount of unutilised court time will 'almost certainly' be at most (1-p)C.

First it is necessary to set a probability value for 'almost certainly'. In the examples in this report we have used a probability of 0.95. The proportion p is then determined by solving P(D>pC)=0.95 for p.

Again assuming Z is a standard normal variate, P(D>pC) is equivalent to P(Z>Q) where $Q=(pC-n\tau)/\sigma\sqrt{n}$.

We also know, from statistical tables for the standard normal distribution, that P(Z>-1.645)=0.95. Therefore, solving for P(D>pC)=0.95 for p we get

$$p = [-1.645(\sigma\sqrt{n}) + n\tau]/C$$

More generally, if a different probability, say α, is selected for the degree of certainty (instead of 0.95) then

$$p = [(\sigma\sqrt{n})Q_\alpha + n\tau]/C \tag{2.3}$$

where Q_α is such that $P(Z>Q_\alpha)=\alpha$.

Using our example data we can now determine the amount of unutilised court time for a court capacity of 4735 court days (that is, the capacity we determined was required to dispose of our 1000 trials, with only a 5% risk of demand exceeding capacity). Substituting in (2.3) with $\alpha=0.95$ we get:

$$p = [(9.4856\ \sqrt{1000}) \times (-1.645) + 4242] / 4735 = 0.79$$

We are therefore 95% certain to use 79% of our court capacity. This means up to 21% of court capacity will be unutilised.

OPTIMISING CAPACITY

In practice each jurisdiction should determine what trade-off it is prepared to make between maximising court utilisation and minimising the risk of demand exceeding capacity. We now explore the relationship between these two factors in more detail.

Suppose we wish to be 95% certain of using a specified proportion p of the available court capacity. We want now to examine how the risk of court delay changes as we increase p towards 1.0. Figure 4 shows how we go about this.

We begin by constructing a family of curves like the two drawn in Figure 3, one for each value of p, starting at p=1.0. We then draw a horizontal line across the graph at the 95% level. Where that line intersects each curve we draw a vertical line downwards. The point at which each vertical line intersects the curve for p=1.0 gives the probability of demand exceeding capacity for that proportion of court capacity used with 95% certainty. So, for the example in Figure 4, we can be 95% certain of having less than 5% of court capacity unutilised (when p=0.95 and 1-p=0.05) only if we accept a risk of 84% of not meeting the demand for court time. Alternatively if we are prepared to have up to 20% of court capacity unutilised (with 95% certainty) the risk of not meeting the demand for court time is only 7%.

The points constructed as described above, for a full range of values of p, may then be plotted in a second graph. The result is Figure 5. Each point in this figure corresponds to a different value for court capacity.

Figure 5 exhibits the trade-off between the proportion of court capacity we use and the risk we take that demand for court services will exceed their supply. The figure shows, for the data on which it was based, that if we were to choose a court capacity with the aim of using 90% of that capacity, we would be taking on a 60% risk of not being able to meet the demand for court services. Alternatively, if we were to choose a court capacity such that there was about a 30% risk of demand exceeding capacity, we would only be certain to use 85% of our court capacity.

It is obvious that the greater the court utilisation, the greater the risk of court delay. The practical implication of this is that court administrators anxious to avoid court delay must build spare capacity into their court systems. Clearly, then, the optimum choice of court capacity is dependent on setting an acceptable risk of court delay, on the one hand, and an acceptable court utilisation rate, on the other.

Figure 4: Probability of demand exceeding a proportion p of capacity, as a function of capacity

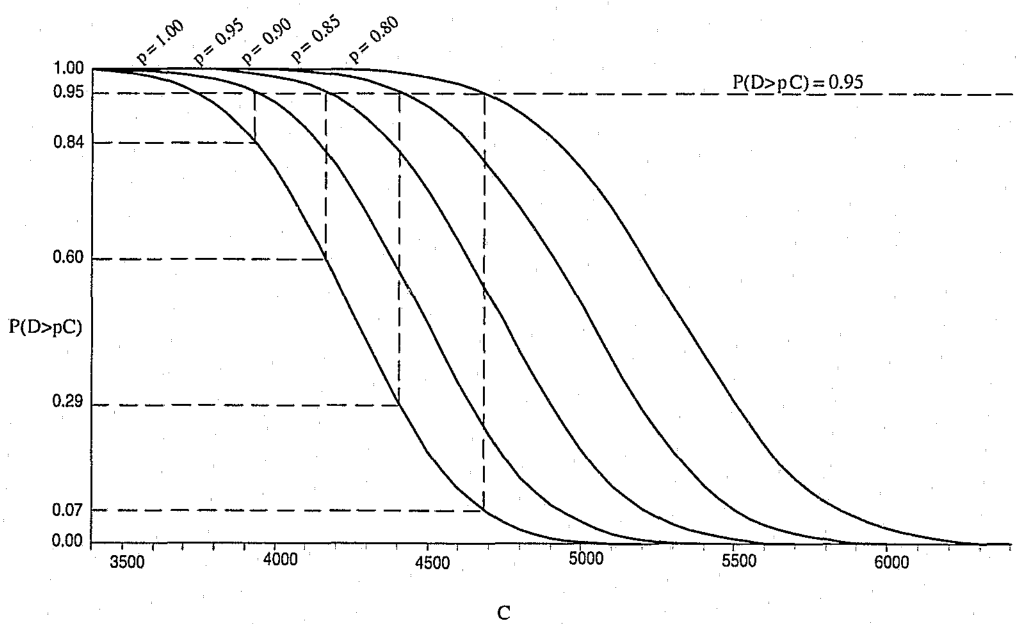
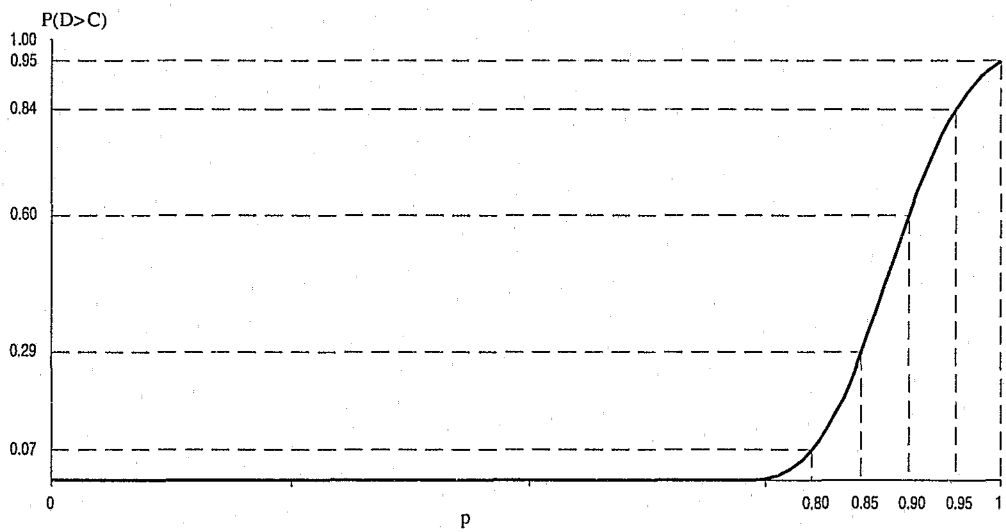


Figure 5: Probability of demand exceeding capacity, as a function of p, where p is chosen such that the probability of demand exceeding a proportion p of capacity is 0.95



ESTIMATING THE SIZE OF THE BACKLOG

In some circumstances we may wish to estimate the possible backlog of cases which may result if court capacity is not increased. We do this by first of all calculating the number of cases which *can* be disposed of with the existing court capacity. If this number is less than the number of cases we *need* to dispose of then the difference between the two is an estimate of the resulting backlog.

The problem then is to determine, for a given capacity, C, the number of cases, n, which can 'almost certainly' be disposed of. Again it is necessary to specify a probability value, say α, for 'almost certainly'. Then we can define the problem as determining the number of cases for which the probability of demand *not* exceeding capacity is α. In other words the risk of demand exceeding capacity is to be $1-\alpha$. The problem is then to solve $P(D>C)=1-\alpha$ for n.

Now $P(D>C)=P(Z>Q)$ where Z is a standard normal variate and $Q=(C-n\tau)/\sigma\sqrt{n}$. We therefore need to solve $(C-n\tau)/\sigma\sqrt{n}=Q$ for n where Q is such that $P(Z>Q)=1-\alpha$. The solution is as follows:

$$C-n\tau = Q(\sigma\sqrt{n})$$

Squaring both sides gives

$$C^2 - 2n\tau C + n^2\tau^2 = Q^2\sigma^2 n$$

This is equivalent to

$$\tau^2 n^2 - (2\tau C + Q^2\sigma^2)n + C^2 = 0$$

We can now solve the quadratic equation in n, using the fact that the solution to the general quadratic $ax^2 + bx + c = 0$ is $x=[-b \pm\sqrt{(b^2 - 4ac)}]/2a$.

In terms of our problem then

substituting n for x
τ^2 for a
$-(2\tau C + Q^2\sigma^2)$ for b
and C^2 for c

we get
$$n = \frac{(2\tau C + Q^2\sigma^2) \pm\sqrt{\{ (2\tau C + Q^2\sigma^2)^2 - 4\tau^2C^2 \}}}{2\tau^2}$$

$$n = \frac{(2\tau C + Q^2\sigma^2) \pm\sqrt{\{4\tau CQ^2\sigma^2 + (Q^2\sigma^2)^2\}}}{2\tau^2}$$

$$n = \frac{(2\tau C + Q^2\sigma^2) \pm\sqrt{\{Q^2\sigma^2(4\tau C + Q^2\sigma^2)\}}}{2\tau^2} \qquad (2.4)$$

Note that there are two solutions to the quadratic equation, depending on whether the second term is added or subtracted. Also note that $Q^2=(-Q)^2$ and only terms in Q^2 appear in (2.4). Therefore this expression is also the solution to

$$C-n\tau = -Q(\sigma\sqrt{n})$$

as well as $C - n\tau = Q(\sigma\sqrt{n})$

That is, the two roots of the quadratic equation in (2.4) give values for n for which (i) $P(Z>Q)=1-\alpha$ and (ii) $P(Z>-Q)=1-\alpha$. The larger value of n is the solution for $P(Z>-Q)$ and the smaller value of n is the solution for $P(Z>Q)$.

The value of n for which $P(D>C)$ is therefore:

$$n = \frac{(2\tau C + Q^2\sigma^2) - \sqrt{\{Q^2\sigma^2(4\tau C + Q^2\sigma^2)\}}}{2\tau^2} \qquad (2.5)$$

The solution is then found by substituting in (2.5) the value of Q for which $P(Z>Q)=1-\alpha$.

We can illustrate the use of formula (2.5) by again substituting our sample data. Substituting $\tau=4.2421$, $\sigma=9.4856$, $C=4735$ and $Q=1.645$ (remembering that $P(Z>1.645)=0.05$) we get:

$$n = (40416.2 - 4429.6) / 36.0 = 1000$$

In other words we can be 95% certain of disposing of 1000 trials. Note that this answer is as expected because the value of C=4735 is the capacity we determined was required to dispose of 1000 trials.

If, on the other hand, court capacity was only 4500 court days we would get n=948 and we would therefore be 95% certain of disposing of only 948 trials, that is, 52 less than the 1000 trials we need to dispose of. There would then be a 95% probability of developing a backlog of up to 52 trials.

3. ALTERNATIVES TO CHANGING CAPACITY

We have so far only been concerned with meeting the demand for court time by increasing court capacity. An alternative approach is to reduce demand.

The decision to increase court capacity should not be based on an uncritical acceptance of the level of demand for court services found in a given jurisdiction. The reason for this is that, as every court administrator knows, changes to court capacity are not the only way a government can meet the problem of court delay. There are a variety of ways in which the demand for court services can be influenced either by legislative or administrative means.

The duration of a trial, for example, can be reduced by changing the legislative constraints governing trial procedure (for example, by removing the need for summing-up in short trials). The demand for trials can also be reduced by changing the rules governing the division between summary and indictable offences to ensure that a larger proportion of cases are dealt with on a summary basis or by creating incentives for defendants to plead guilty. Both the number and the duration of trials can be altered by changing the charging practices of prosecuting agencies.

We chose in Section 2 to assume that the case hearing times t_i are identically and independently distributed with mean τ and variance σ^2. To assess the effect of reductions in demand we need to consider the distribution of the t_i in more detail.

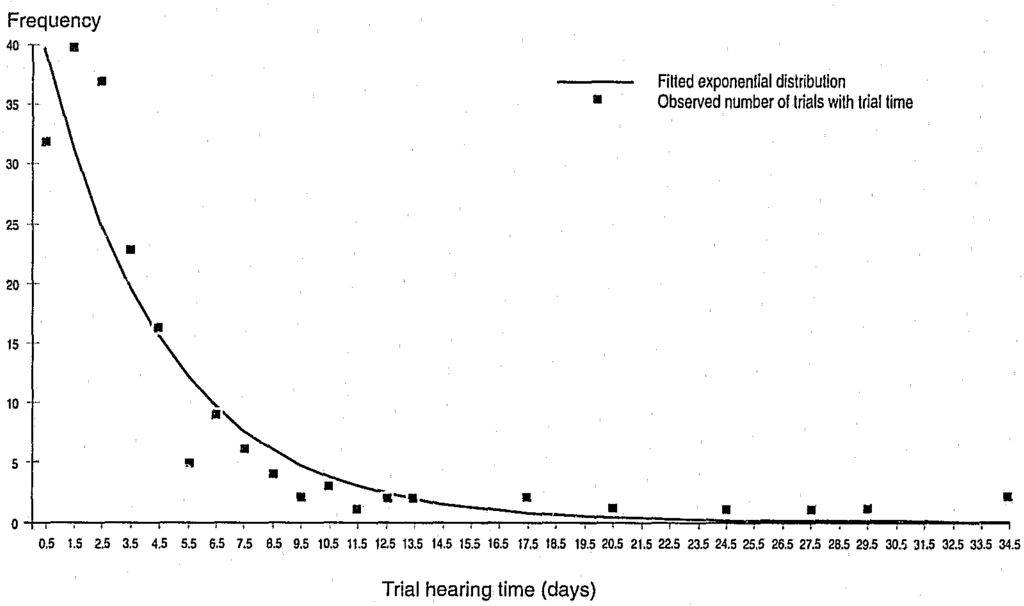
Suppose that, at any point in time for a court case in progress, (i) there is a probability that the court case will come to an end, and (ii) this probability is not dependent on the length of time the court case has already been running. Under this assumption the distribution of case hearing times would be exponential.[3]

Figure 6 shows the distribution of trial hearing times for our sample of 190 trial cases drawn from the Sydney District Criminal Court during 1989.

The trial duration distribution looks exponential. In fact the exponential distribution gives a reasonable fit to the data except for the fact that there are one or two trials whose hearing times are rather longer than would be expected if this were the case. Ignoring this discrepancy for the moment, assume that trial hearing times are exponentially distributed, that is assume that for each t_i the probability distribution is:

$$f(t) = ae^{-at}$$

Figure 6: Frequency distribution of trial hearing times for sample from Sydney District Criminal Court



This distribution has mean $1/a$ and variance $1/a^2$. Note therefore that as the mean changes, so will the variance.

If we proceed on the basis that the component trial time distributions are exponential the demand for trial time can be shown to be normally distributed with mean $n\tau$ and variance $n\tau^2$ where $\tau=1/a$.

The sample data shown in Figure 6 provided estimates for the mean and variance as follows:

$$t = 4.2421 \qquad s^2 = 29.8114 = (5.4600)^2$$

Because the exponential distribution does not fit well in the upper tail of the distribution in Figure 6, the variance of D may actually be larger than $n\tau^2$. (Note that $s^2=1.66t^2$.)

Let us therefore now assume that the demand for court services, D, is normally distributed with mean $n\tau$ and variance $kn\tau^2$ where k is greater than one. The value of such an assumption is that we can explore changes in the mean hearing time without needing a separate estimate of the variance of D, because the variance is a function of the mean.

Note that the example data used in Section 2 had a mean and variance for D, of 4242 and 89977, respectively. This variance is equal to $kn\tau^2$ where k=5, n=1000 and τ=4.2421. A larger variance than the estimate obtained from the sample was chosen deliberately, to demonstrate the trade-off between the proportion of unutilised court time and the probability of demand exceeding capacity. With a smaller variance for D it is often possible to choose a capacity such that both of these are at acceptable levels. Later in this section of the report we shall examine the sensitivity of P(D>C) to the variance of D.

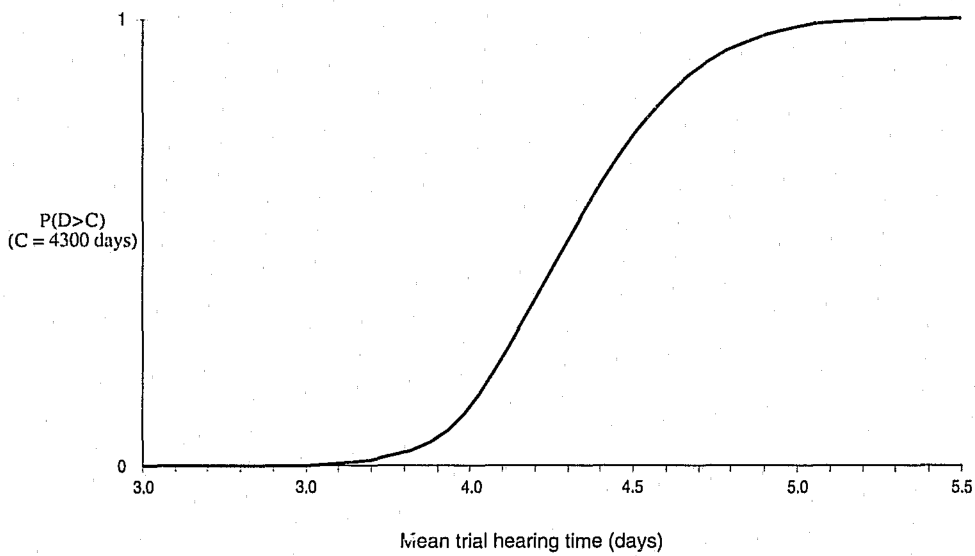
First we will examine the effect of changes in the factors which determine demand for court time.

EFFECT OF CHANGES IN AVERAGE HEARING TIME OR THE NUMBER OF CASES

The average demand for court time, it will be recalled, is given by $n\tau$, where n is the number of cases to be disposed of and τ gives the average duration of a court case. The effect of changes to the average case hearing time, τ, or the number of cases which have to be disposed of, n, can be quite significant. Figures 7 and 8 have been constructed for a fixed value of C=4300 court days. Figure 7 shows the relationship between P(D>C) and τ, the average hearing time, for a fixed value of n=1000. Figure 8 shows the relationship between P(D>C) and n, the number of cases required to be disposed of, for a fixed value of τ=4.2421. Note that in both cases P(D>C) has been determined by reading values for P(Z>Q) from statistical tables for the standard normal distributionwhere $Q=(C-n\tau)/\tau\sqrt{(kn)}$, substituting 4300 for C, and varying values for τ (for Figure 7) and for n (for Figure 8).

It is obvious that, in this example, P(D>C) is extremely sensitive to the average duration of a trial when that duration rises above about 4 days. At 4.1 days the risk in a given year that demand for court services will exceed court capacity is 25%. By the time the average duration of a trial reaches 4.5 days, that risk has risen to nearly 75%. Equally, a fall from a 75% risk that demand will exceed capacity to a 25% risk, requires a decrease of less than 100 cases per annum.

Figure 7: Probability of demand exceeding capacity, as a function of trial hearing time



Clearly reductions in either the average duration of trials or the number of trials have the potential to reduce substantially the overall demand for trial court services. This means that, in planning court services it is important to explore means of reducing the demand for services as alternatives to increasing the court capacity of the court system.

EFFECT OF CHANGES IN THE VARIANCE OF HEARING TIMES

Figure 9 plots, as does Figure 7, P(D>C) against the average duration of a trial for a fixed court capacity of C=4300 days and a fixed value of n=1000 trials. The three curves, however, show this relationship for three different values of the variance of trial times. They are, respectively, twice, five times and ten times the variance one would expect if the underlying distribution of hearing times was exponential. It can be seen that as the variance of hearing times is reduced, the steepness of the curve increases. From this we may infer that reducing the variance of trial times increases court utilisation for a given level of risk that demand for court services will exceed capacity.

For example, for a level of risk of 30% that demand will exceed capacity, Figure 5 shows that we will use 85% of court capacity (with 95% certainty). In constructing Figure 5 we used a value of k=5, that is, five times the variance expected if the underlying distribution of hearing times was exponential. If, instead, we were to use a value of k=2 to construct Figure 5

Figure 8: Probability of demand exceeding capacity, as a function of the number of trials

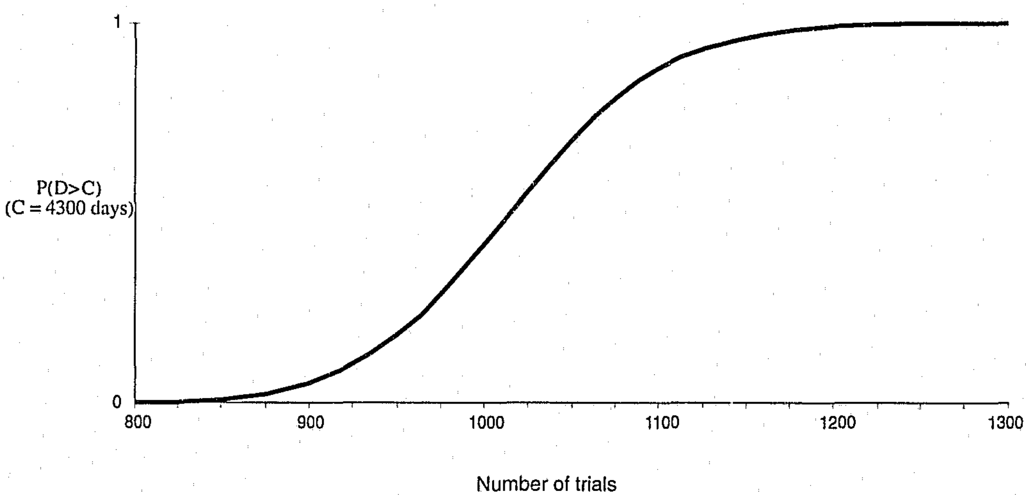
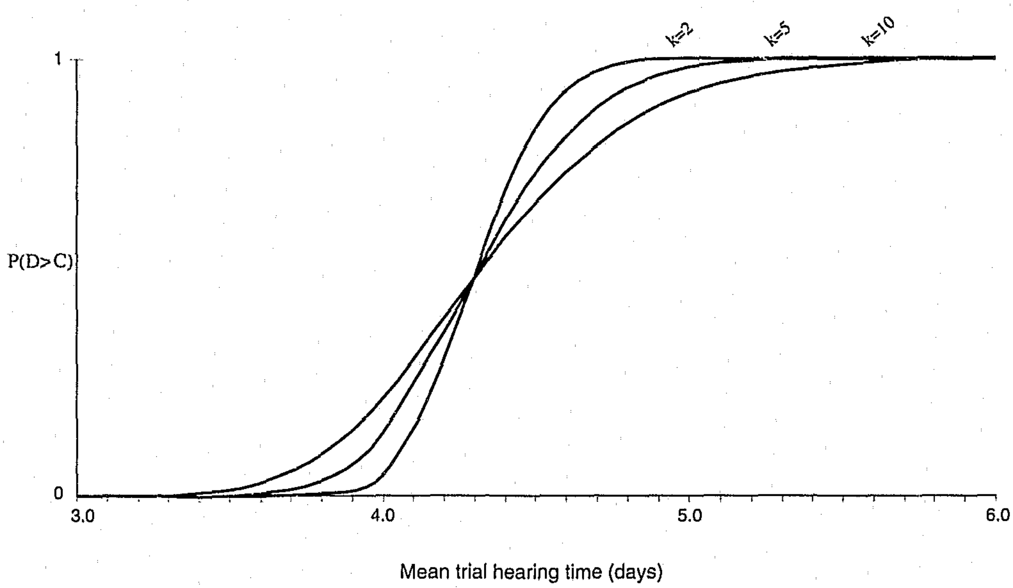


Figure 9: Probability of demand exceeding capacity, as a function of trial hearing time



we would find that for a 30% risk of exceeding capacity we would use 90% of capacity (with 95% certainty).

Reducing the variance of trial times therefore ensures less unutilised court time, for a given risk of demand exceeding capacity. Practical applications of this principle include, for example, planning capacity separately for short and long matters.

4. PRACTICAL APPLICATION OF THE METHOD

There are essentially two steps involved in planning court capacity using the method described in Section 2. Step one is to measure the demand for court time. Step two is to calculate optimum court capacity given that level of demand. This section of the report describes each of these steps in turn and gives an illustration of the use of the method.

DETERMINING DEMAND FOR COURT TIME

There are two factors which jointly determine the demand for court services in a particular jurisdiction. The first of these factors is n, the number of cases to be disposed of. The second factor is the amount of court time consumed by those cases. For this we need to measure $f(t)$, the frequency distribution for court time consumed per case.

Measuring the number of cases to be disposed of

The number of cases registered provides the starting point for the measurement of n. However not all cases registered will eventually require court time. This is especially true of trials. Typically, a substantial proportion of matters registered as trials will never go on as trials. This happens either because the accused changes plea or absconds or has the charges against him or her dropped or, occasionally, because he or she dies. For example, in the NSW District Court, trial attrition due to these various circumstances is such that less than 50% of trials registered actually go on as trials. The magnitude of trial attrition, however, is likely to vary both between jurisdictions and within a jurisdiction over time. This suggests that the number of matters registered as trials which eventually go on as trials ought, ideally, to constitute a regular feature of a jurisdiction's management statistical output.

Measurement of the number of cases to be disposed of involves taking the number of cases registered and reducing it by the proportion of those cases which do not eventually require court time. In determining the latter proportion it is important to include only those matters registered in a particular year which do not require court time either in that year *or in any succeeding year*. Obviously in a particular jursidiction there may be a very small residue of cases whose status remains uncertain even after several years. This raises the question of over what period one should endeavour to calculate a value of n. The choice is somewhat arbitrary but, generally speaking, a waiting period should be chosen which is sufficient to ensure that the status of at least 95% of the matters originally registered is fully resolved.

Measuring the hearing time distribution

Practical determination of the distribution of the hearing time for cases requires considerable care, both because it is a very potent influence on the overall demand for court services and because it is difficult to measure accurately. The nominal starting point for constructing the hearing time distribution, f(t), is the period which elapses between case commencement and case completion. In many instances this can often be determined from case files simply by taking the difference between the date of case commencement and the date of case completion, making due allowance for weekends.[4] Note, however, that for convicted offenders the date of case completion must be taken as the date on which sentencing of the offender is completed (though if there is an adjournment between conviction and sentence the period of the adjournment should be removed). Trial hearing time should also include any court time consumed by a case during the various interlocutory hearings in the lead-up to the trial hearing itself. If these cannot be empirically determined they would have to be 'guestimated' by experienced staff.

Apart from taking all reasonable steps to secure as accurate a measure as possible of the amount of court time consumed by any particular case, the most important precondition in successfully determining the distribution of demand for court services is ensuring an adequate and representative sample of hearing duration measurements. The issue of representativeness is of particular importance because in most jurisdictions some courts specialise in long matters, while others deal mainly in short matters (for example, sentence hearings, bail applications etc.). The method of choosing courts from which to sample case hearing durations for measurement should ensure that all cases are equally likely to be chosen. Where, in a particular jurisdiction, courts specialise in long or short trials, therefore, the sample sizes from each type of court should be chosen so as to reflect the relative proportions of short and long trials in the population of trials.[5]

Within these constraints, generally speaking, a sample of 300 measurements should be sufficient to determine the distribution of the hearing times. The mean t and variance s^2 of the hearing times can then be calculated from equations (4.1) and (4.2) below (where m is the sample size and t_i is the hearing time for case i, for values of i from 1 to m):

$$t = \Sigma t_i / m \qquad (4.1)$$

$$s^2 = [\Sigma t_i^2 - (\Sigma t_i)^2/m] \ / \ (m - 1) \qquad (4.2)$$

DETERMINING OPTIMUM COURT CAPACITY

Our measure of court capacity is in court days. Theoretically actual court capacity for a given year is simply the product of the number of days per week each court sits, the annual number of court sitting weeks and the number of courts. Thus if there are 21 courts each sitting five days per week for 41 weeks of the year, the court capacity for the year is 4305 days. In practice, however, some of this capacity is inevitably lost for a variety of different reasons. For example, trials may end early during a day but too late for the court to be used to begin another trial on that day. Trials can be adjourned or aborted. Courts can be also be temporarily closed due to misadventure or industrial problems.

Ideally the proportion of lost court time should be empirically determined and used to adjust the estimate of optimum court capacity. The way this is done is explained below.

Since court delay results whenever demand exceeds capacity and one object of court management is to reduce unnecessary delay, the first step in choosing the optimum trial court capacity is to set a tolerable risk that demand will exceed capacity. We shall assume that the aim is to maintain the risk that demand will exceed capacity at or below a specified probability, r. Given this assumption the required court capacity, C, can be found using equation (2.2) from Section 2:

$$C = (\sigma\sqrt{n})Q_r + n\tau \qquad (4.3)$$

where (1) τ and σ^2 are, respectively, the mean and variance of the hearing times and are estimated by t and s^2 above;

(2) n is the number of cases to be disposed of;

and (3) Q_r is such that $P(Z>Q_r)=r$, where Z is a standard normal variate.

Expression (4.3) gives us a means of determining the court capacity required to dispose of n cases for a specified risk, r, of demand exceeding capacity. However as noted above some court time is inevitably lost. If the proportion of lost time is known to be L, and the *actual* court capacity that needs to be planned, in order to provide an amount C of court time is denoted by C_a then we can determine C_a as follows. We know that an amount LC_a of court time is lost and therefore that only an amount $(1 - L) C_a$ of court time is available. The capacity that must be planned for is therefore given by:

$$C_a = C / (1 - L)$$

We found in Section 2, that the lower the risk that demand exceeds capacity, the higher the risk of unutilised court time. Since the amount of unutilised court time is one measure of the economy of service, we are also interested in the amount of court time we are 'certain' to use. We need first to set a probability value for the level of certainty. In order to determine the percentage of court capacity, p, which will be utilised with probability α, we use equation (2.3) from Section 2:

$$p = [(\sigma\sqrt{n})Q_\alpha + n\tau]/C \tag{4.4}$$

where Q_α is such that $P(Z>Q_\alpha)=\alpha$.

To determine an optimum capacity it may be necessary to calculate several values of C, and the corresponding values for p, by varying r, the risk of demand exceeding capacity. Each administrator can then determine the risk of demand exceeding capacity he or she is prepared to take in order maximise court utilisation.

ESTIMATING THE POSSIBLE BACKLOG OF CASES

In many situations we are also going to be interested in knowing how many trials (N) a given level of capacity can dispose of with a specified probability, say α. (Typically we would set a relatively high value for α, say 95%.) This calculation would allow us to contrast the number of cases which have to be disposed of with the number of cases our court system is capable of handling. To obtain the number of trials a court system of capacity C can dispose of with probability α we apply equation (2.5) derived from Section 2:

$$N = \frac{(2\tau C + Q^2\sigma^2) - \sqrt{\{Q^2\sigma^2(4\tau C + Q^2\sigma^2)\}}}{2\tau^2} \tag{4.5}$$

where Q is such that $P(Z>Q)=1-\alpha$.

A WORKED EXAMPLE

Suppose the number of matters expected to be registered as trials in a specified year is 3360. For the sake of argument we will assume that there are no pending cases.[6] From past experience it is known that only about 40% of matters registered as trials will ever go on as trials. The number of trials to be disposed of is therefore 3360 x 0.4=1344.

There are presently 21 trial courts operating at 5 days per week for 41 weeks of the year. Therefore the current capacity of the court system is 4305 days.

Assume also that it is known that 2% of court time is lost so that only 98% of these 4305 court days, or 4219 court days, are available for hearing cases.

From a sample of trials we have estimates of the mean and variance of hearing times as follows:

$t = 4.2421$ $\qquad\qquad$ $s^2 = 29.8114 = (5.4600)^2$

What court capacity would be required to dispose of 1344 trials?

To calculate the required court capacity we must first specify a level of risk of demand exceeding capacity. Suppose we wish to take only a 5% chance that demand will exceed capacity. Then we apply formula (4.3), substituting 1344 for n, 4.2421 for τ, 5.4600 for σ and 1.645 for Q_r (since $P(Z>1.645)=0.05$):

$$C = (\sigma\sqrt{n})Q_r + n\tau$$
$$= 5.4600 \times \sqrt{(1344)} \times 1.645 + 1344 \times 4.2421$$
$$= 6031$$

Thus a court capacity of 6031 court days is required to dispose of 1344 cases with a 5% risk of demand exceeding capacity. Note, however, that because 2% of court capacity is lost time, the actual capacity which must be planned in order to provide 6031 court days for hearing trials, is 6031 / 0.98=6154 court days.

What proportion of the court capacity is 95% certain to be utilised? Noting that $P(Z>-1.645)=0.95$ and substituting into equation (4.4) values as above for n, τ and σ, and 6031 for C, we obtain:

$$p = [(\sigma\sqrt{n})Q_\alpha + n\tau]/C$$
$$= [5.4600 \times \sqrt{(1344)} \times -1.645 + 1344 \times 4.2421] / 6031$$
$$= 0.89$$

Therefore to dispose of 1344 trials we need a court capacity of 6031 court days to ensure that the risk of demand exceeding capacity is 5% and we can be 95% certain that the maximum unutilised court capacity is 11%.

What additional court utilisation could we achieve if we were prepared to increase the risk of demand exceeding capacity to 10%?

Again applying formula (4.3) with r=0.10 and therefore Q_r=1.282 we get

$$C = 5.4600 \times \sqrt{(1344)} \times 1.282 + 1344 \times 4.2421 = 5958$$

Substitution in (4.4) with C=5958 gives

$$p = [5.4600 \times \sqrt{(1344)} \times -1.645 + 1344 \times 4.2421] / 5958$$
$$= 0.90$$

Therefore with a court capacity of 5958 court days we can be 95% certain that the maximum unutilised court capacity is 10%. However with this reduced capacity there would be a 10% risk of demand exceeding capacity.

Further calculations show that to increase court utilisation to 95% requires an unacceptably high risk of demand exceeding capacity.

What number of trials are we 95% certain to dispose of, given the existing capacity of 4305 days?

To obtain this information we apply equation (4.5) with values for τ and σ as before and with C=4219 and Q=1.645 (since P(Z>1.645=0.05=1-0.95).

Note first that $Q^2\sigma^2 = (1.645 \times 5.4600)^2 = 80.6709$

Then we obtain:

$$N = \frac{(2tC + Q^2\sigma^2) - \sqrt{\{Q^2\sigma^2(4\tau C + Q^2\sigma^2)\}}}{2\tau^2}$$

$$= \frac{(2 \times 4.2421 \times 4219 + 80.6709) - \sqrt{\{80.6709\,(4 \times 4.2421 \times 4219 + 80.6709)\}}}{[2 \times (4.2421)^2]}$$

$$= 930$$

Given, therefore, that we have to dispose of 1344 trials we can expect a shortfall or backlog of up to 1344-930=414 trials.

NOTES

[1] This obviously takes no account of losses due to early adjournments or when the court is occupied for non-trial purposes. We shall address this issue in Section 4.

[2] See, for example, Feller, W., *An Introduction to Probability Theory and Its Applications*, 2nd edn 1957, John Wiley and Sons, p. 239.

[3] See, for example, Feller, op. cit., p. 412.

[4] If this approach is adopted it would be advisable to add half a day to the result. The correction is important because if a hearing began and ended on the same day, the difference between the two dates in question would be erroneously judged to be zero. The addition of half a day, though somewhat arbitrary, reduces the trial hearing time measurement error for trials beginning or ending at some point during the day. Ideally, of course, one should conduct an empirical study on the amount of court time consumed by a sample of trials.

[5] Note, however, that where the issue of trial court capacity arises only in respect (say) of courts attached to a particular registry, sampling should take place only among trials disposed of by courts attached to that registry.

[6] If there were, some or all of these would have to be added to the cases which arrive.

APPENDIX: PROGRAMS FOR USE ON AN HP 32S CALCULATOR

The calculations required to carry out the method described in this report can very easily be programmed for a programmable calculator.

We used a Hewlett Packard 32S scientific calculator with Reverse Polish Notation. The programs used and instructions for their use are provided below for the information of those who may wish to use them.

PROGRAMS

Program C

C01	LBL C
C02	ENTER
C03	STO N
C04	SQRT
C05	RCL S
C06	x
C07	RCL Q
C08	x
C09	RCL N
C10	ENTER
C11	RCL T
C12	x
C13	÷
C14	RTN

Program P

P01	LBL P
P02	ENTER
P03	STO C
P04	RCL N
P05	SQRT
P06	RCL S
P07	x
P08	-1.645
P09	x
P10	RCL N
P11	ENTER
P12	RCL T
P13	x
P14	+
P15	RCL C
P16	÷
P17	RTN

Program N

N01	LBL N
N02	ENTER
N03	2
N04	x
N05	RCL T
N06	x
N07	STO Z
N08	RCL Q
N09	ENTER
N10	RCL S
N11	x
N12	x^2
N13	STO W
N14	+
N15	RCL Z
N16	ENTER
N17	2
N18	x
N19	RCL W
N20	+
N21	RCL W
N22	x
N23	SQRT
N24	-
N25	2
N26	÷
N27	RCL T
N28	x^2
N29	÷
N30	RTN

PROGRAM C

Program C calculates C using formula (4.3), that is, it calculates the capacity required to dispose of n cases. To use the program it is necessary to store the value for t in Register T, the value for s in Register S and the value for Q in Register Q. (Note that t and s^2 are respectively the estimates of the mean and variance of the hearing times and Q is such that P(Z>Q)=r where r is the specified risk of demand exceeding capacity.) Then the value of n is used to enter the program:

(value of n) XEQ C

For example, for the worked example shown on page 23, we would store the parameter values in the appropriate registers as follows:

4.2421 STO T

5.4600 STO S

1.645 STO Q

then entering 1344 XEQ C gives the result 6030.6570.

Program C also stores the value for n in Register N and this value is used in Program P.

PROGRAM P

Program P calculates p using formula (4.4) with Q_{α}=-1.645, that is, it calculates the proportion of court capacity used with 95% probability. The program can be used immediately after obtaining a value for C, from Program C, by entering XEQ P while the value of C is still displayed. For example, 6030.6570 XEQ P gives the result 0.8908 (provided no changes to the values stored in Registers T, S, Q and N have been made).

Otherwise, to use Program P it is necessary to store the value for n in Register N, the value for t in Register T, the value for s in Register S and the value for Q in Register Q. (Note that n, t and s are as above and Q is such that P(Z>Q)=α where α is the probability that demand exceeds pC.) Then the value of C should be entered, followed by XEQ P.

Note that a fixed value of Q_{α}=-1.645 is used in Program P, that is, the program calculates p such that P(D>pC)=0.95. If an alternative degree of certainty is required the program can be modified. Statement P08 can be replaced with another value for Q, or with the statement RCL A, in which case the value for Q_{α} would need to be stored in Register A prior to program execution.

USING PROGRAMS C AND P TOGETHER

The following example shows how these two programs can be used to determine an optimum capacity.

The values 4.2421 and 5.4600 should be stored in Registers T and S, respectively.

The number of trials to be disposed of is 1344.

For a 5% risk of demand exceeding capacity Q=1.645:

Entering	1.645 STO Q		
	1344 XEQ C	gives	C = 6030.6570
	XEQ P	gives	p = 0.8908

For a 10% risk of demand exceeding capacity Q=1.282:

Entering	1.282 STO Q		
	1344 XEQ C	gives	C = 5957.9964
	XEQ P	gives	p = 0.9017

For a 15% risk of demand exceeding capacity Q=1.036:

Entering	1.036 STO Q		
	1344 XEQ C	gives	C = 5908.7553
	XEQ P	gives	p = 0.9092

For a 20% risk of demand exceeding capacity Q=0.842:

Entering	0.842 STO Q		
	1344 XEQ C	gives	C = 5869.9229
	XEQ P	gives	p = 0.9152

For a 50% risk of demand exceeding capacity Q=0.000:

Entering	0.000 STO Q		
	1344 XEQ C	gives	C = 5701.3824
	XEQ P	gives	p = 0.9422

Under these circumstances it can be seen that to ensure 94% of court capacity is utilised requires taking a 50% risk of demand exceeding capacity. Alternatively, for a reasonable risk of demand exceeding capacity of say 10%, we can only be certain of using 90% of court capacity.

PROGRAM N

Program N calculates N using formula (4.5), that is, it calculates the number of cases a court with capacity C can dispose of with a specified probability α. To use the program it is necessary to store the value for t in Register T, the value for s in Register S and the value for Q in Register Q. (Note that t and s are as above and Q is such that $P(Z>Q)=1-\alpha$.) To run the program a value for C should be entered, followed by XEQ N.

For example, with 4.2421 stored in Register T, 5.4600 stored in Register S and 1.645 stored in Register Q, entering 4219 XEQ N gives the result 929.9868, as in the worked example on page 24.

Note also that Program N uses Registers Z and W as working storage during program operation. Any data stored in these registers will therefore be lost during program execution.