4
Consecutive follow-up intervals

In the last chapter we touched on the difficulty of estimating the probability of failure during a fixed follow-up period when the observation times for some subjects are censored. A second problem with fixed follow-up periods is that it may be difficult to compare the results from different studies; a five-year probability of failure can only be compared with other five-year probabilities of failure, and so on. Finally, by ignoring *when* the failures took place, all information about possible changes in the probability of failure during follow-up is lost.

The way round these difficulties is to break down the total follow-up period into a number of shorter consecutive intervals of time. We shall refer to these intervals of time as *bands*. The experience of the cohort during each of these bands can then be used to build up the experience over any desired period of time. This is known as the *life table* or *actuarial* method. Instead of a single binary probability model there is now a sequence of binary models, one for each band. This sequence can be represented by a conditional probability tree.

4.1 A sequence of binary models

Consider an example in which a three-year follow-up interval has been divided into three one-year bands. The experience of a subject during the three years may now be described by a sequence of binary probability models, one for each year, as shown by the probability tree in Fig.4.1. The four possible outcomes for this subject, corresponding to the tips of the tree, are

1. failure during the first year;
2. failure during the second year;
3. failure during the third year;
4. survival for the full three-year period.

The parameter of the first binary model in the sequence is π^1, the probability of failure during the first year; the parameter of the second binary model is π^2, the probability of failure during the second year, given the subject has not failed before the start of this year, and so on. These are

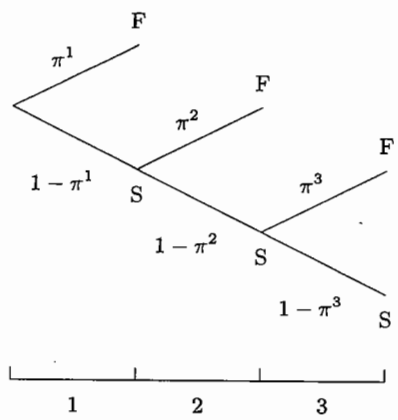


Fig. 4.1. A sequence of binary probability models.

all conditional probabilities — conditional on not having failed before the start of the year in question. The reason the probabilities are written with superscripts is that we have adopted the convention that a superscript is used to index *time*, and a subscript is used to index subjects or groups of subjects. It is important to distinguish these two situations, and using subscripts for both can be confusing.*

Suppose, for illustration, that the probability of failure is 0.3 in the first year; 0.2 in the second year, given the subject survives the first year without failure; and 0.1 in the third year, given the subject survives the first two years without failure. These illustrative values for the three conditional probabilities are shown on the conditional probability tree in Fig.4.2.

In this tree, the four final outcomes listed above correspond to the tips of the tree, and their probabilities can be calculated by multiplying conditional probabilities along the branches of the tree in the usual way. For example, the probability of the second outcome is made up from the probability that the subject survives the first year (0.7), multiplied by the probability that the subject fails during the second year (0.2). Using this rule, the four possible outcomes for any subject occur with probabilities:

$$0.3$$
$$0.7 \times 0.2$$
$$0.7 \times 0.8 \times 0.1$$
$$0.7 \times 0.8 \times 0.9$$

*Note that π^2 does not refer to $\pi \times \pi$. To avoid confusion we shall always use brackets when taking powers; for example, the square of π will be written $(\pi)^2$.

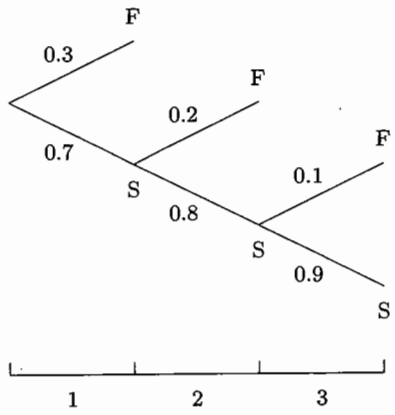


Fig. 4.2. Illustrative values for the conditional probabilities.

These probabilities work out to be 0.3, 0.14, 0.056, and 0.504, and these add to 1, as they should, since there are no other possible outcomes. The probability of failing at *some* stage is

$$0.3 + 0.14 + 0.056 = 0.496.$$

More conveniently this probability can be found by subtracting from 1 the probability of surviving the three years without failing, giving

$$1 - 0.504 = 0.496.$$

The probabilities of surviving one, two, and three years without failing are called the *cumulative survival probabilities* for the cohort. They are calculated by multiplying the conditional probabilities of surviving each year, and in this case are:

$$0.7$$
$$0.7 \times 0.8$$
$$0.7 \times 0.8 \times 0.9.$$

which work out to be 0.7, 0.56, and 0.504.

Exercise 4.1. In a three-year follow-up study the conditional probabilities of failure during the first, second, and third years are 0.05, 0.09, and 0.12 respectively. Draw a probability tree for the possible outcomes for a new subject, and label the branches of the tree with the appropriate conditional probabilities. Calculate the probability of each of the outcomes, and the probabilities of surviving

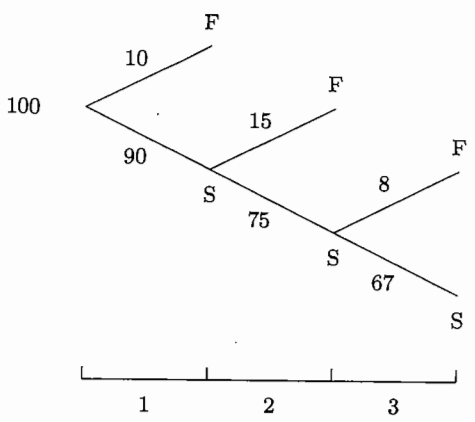


Fig. 4.3. Survival of 100 subjects through three time bands.

one, two, and three years without failing. Calculate also the probability of failing at some time during the three-year follow-up.

4.2 Estimating the conditional probabilities of failure

Suppose that 100 subjects join the cohort at the start of the three-year interval and that 10 fail during the first year, 15 during the second, and 8 during the third, leaving 67 who survive until the end of three years (see Fig.4.3). Assuming the same conditional probabilities of failure for each of the 100 subjects, these data can be used to estimate their most likely values.

Intuitively it seems sensible to use the experience of those subjects who are observed in each year to estimate the conditional probability of failure during that year. The most likely values of the three conditional probabilities would then be

$$\frac{10}{100}, \frac{15}{90}, \frac{8}{75},$$

but is this a legitimate thing to do? It corresponds to regarding the three-year follow-up study as equivalent to three separate and independent one-year follow-up studies in which the subjects come from the survivors of the previous year. In fact this is a legitimate thing to do because the likelihood for π^1, π^2, and π^3 is the same whether the data are regarded as coming from one three-year study or from three one-year studies. This may be shown algebraically as follows.

The probabilities of the four possible outcomes in the three-year study are

$$\begin{gathered}\pi^1 \\ (1-\pi^1)\pi^2 \\ (1-\pi^1)(1-\pi^2)\pi^3 \\ (1-\pi^1)(1-\pi^2)(1-\pi^3)\end{gathered}$$

A subject who fails during the first year therefore contributes

$$\log(\pi^1)$$

to the log likelihood. A subject who fails during the second year contributes

$$\log(1-\pi^1)+\log(\pi^2),$$

a subject who fails during the third year contributes

$$\log(1-\pi^1)+\log(1-\pi^2)+\log(\pi^3),$$

and a subject who survives all three years contributes

$$\log(1-\pi^1)+\log(1-\pi^2)+\log(1-\pi^3).$$

Multiplying these by the numbers of subjects with each outcome, that is 10, 15, 8, and 67 respectively, and adding, gives a total log likelihood of

$$\begin{gathered}10\log(\pi^1)+90\log(1-\pi^1) \\ +15\log(\pi^2)+75\log(1-\pi^2) \\ +8\log(\pi^3)+67\log(1-\pi^3).\end{gathered}$$

This is the same as the log likelihood obtained by regarding the data as from three separate and independent one-year studies; the first based on 10 failures and 90 survivors, the second on 15 failures and 75 survivors, and the third on 8 failures and 67 survivors.

Exercise 4.2. If we were to adopt the more restrictive model that π^1, π^2, π^3 are all equal with common value π, what would be the most likely value of π?

This exercise makes it clear that, in the analysis of such studies, the basic atom of data is not the subject, but the observation of one subject through one time band.

4.3 A cohort life table

In cohorts where subjects are examined at yearly intervals, the data are often presented in the form of numbers of failures and censorings occurring each year. An example is given in Table 4.1, which refers to survival of a

Table 4.1. Survival by stage at diagnosis

	Stage I			Stage II		
Year	N	D	L	N	D	L
1	110	5	5	234	24	3
2	100	7	7	207	27	11
3	86	7	7	169	31	9
4	72	3	8	129	17	7
5	61	0	7	105	7	13
6	54	2	10	85	6	6
7	42	3	6	73	5	6
8	33	0	5	62	3	10
9	28	0	4	49	2	13
10	24	1	8	34	4	6

group of women with cancer of the cervix diagnosed at either stage I or stage II. The women are examined annually, and censoring occurs if they cease attending the clinic; N is the number alive and still under observation at the start of each time band, D is the number who die during each band, and L is the number censored during each band.

The estimation of survival experience of the stage I women over the first four years is shown in Fig.4.4. Of the 110 subjects who started the first year, 5 die and 5 are censored. The effective size of the cohort in the first year is taken to be 107.5 and the probability of a subject dying during the first year, given the subject was alive at the start of the year, is estimated to be $5/107.5 = 0.0465$. The conditional probability of surviving the year is estimated to be

$$1 - 0.0465 = 0.9535.$$

The calculations of failure and survival probabilities are shown in Fig.4.4. The cumulative survival probabilities are found by multiplying the conditional survival probabilities for each year. For example, the cumulative probability of surviving 3 years is

$$0.9535 \times 0.9275 \times 0.9152 = 0.8093.$$

Exercise 4.3. Using Table 4.1, draw a tree showing the survival experience for stage II women over the first four years, and calculate the conditional survival probabilities for each of these years.

A table of cumulative survival probabilities by year is called a *life table*, and a plot of the cumulative survival probabilities against years survived is called a *survival curve*. The survival curves for both stage I and stage II

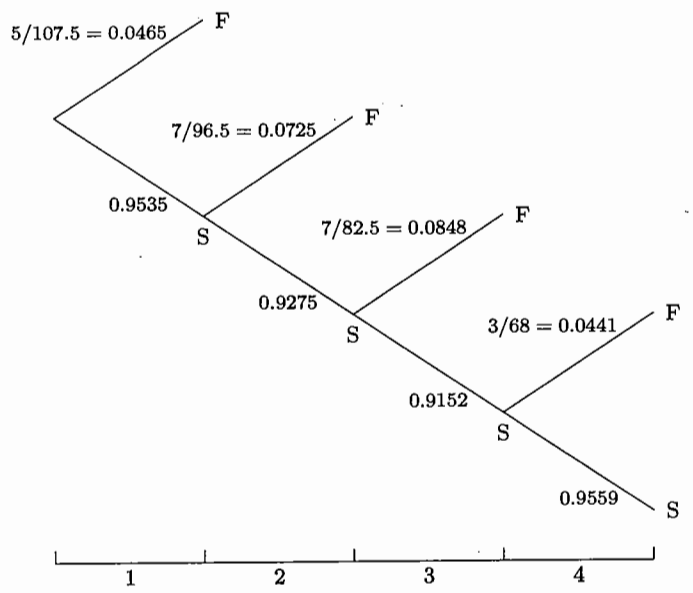


Fig. 4.4. Estimated conditional probabilities for stage 1 women.

women are shown in Fig.4.5. It is conventional to start survival curves at a probability of one for surviving at least zero years. These plots are useful for studying whether the probability of failure is changing with follow-up time, and for calculating survival probabilities for different periods of time.

Exercise 4.4. Use Fig.4.5 to read off the five-year survival probabilities in each of the two groups.

4.4 The use of exact times of failure and censoring

In the calculations described above, the conditional probability of failure during each time band has been estimated by assuming, as in Chapter 3, that half the losses during the band occurred at the start and half at the end. If the individual times at which failure (or censoring) occur are known then it is possible to avoid this assumptions by choosing the bands so short that each failure occupies a band by itself. Such a choice of bands is shown in Fig.4.6 for the early follow-up experience of 50 subjects. The horizontal line represents follow-up time, failures are marked as •, and losses as ×. The bands are shown by vertical bars. Only the first few events are shown.

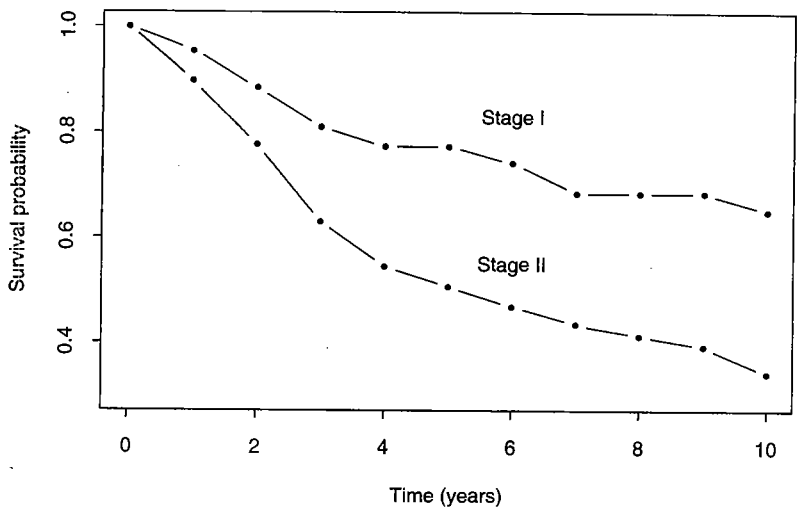


Fig. 4.5. Survival curves for Stage I and Stage II women.

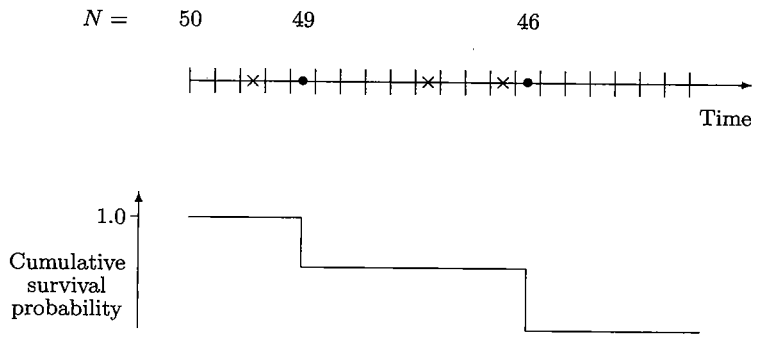


Fig. 4.6. Early follow-up experience of 50 subjects.

For bands in which there are no failures the estimated survival probability is 1. For bands which contain a failure the estimated survival probability is $1 - 1/N$ where N is the number at risk just before the failure. Thus for the band which contains the first failure $N = 49$ and the estimated survival probability is $1 - 1/49 = 48/49$. The estimate of the cumulative survival probability up to the end of this band is

$$1 \times 1 \times \cdots \times 48/49 = 0.9796.$$

For the band which contains the second failure $N = 46$, so the estimated survival probability for this band is $1-1/46 = 45/46$. The cumulative probability of survival up to the end of the fourth band is therefore estimated at

$$1 \times \cdots \times 48/49 \times 1 \times \cdots \times 45/46 = 0.9583.$$

These calculations continue until there are no more bands which contain failures.

The bands containing each failure can be made so short that they refer to the actual time of failure. When this is done the cumulative survival probability over time takes the value 1 until the first failure, when it drops to 0.9796; then it stays at 0.9796 until the second failure when it drops to 0.9583, and so on. The plot of cumulative survival probability versus time survived takes the stepped shape shown in Fig.4.6, where the steps occur at the failure times.

This method of estimating the cumulative survival probabilities is called the *Kaplan–Meier* method, after the authors of the paper which showed that this procedure yields the most likely value of the survival curve. It is widely used in clinical follow-up studies for which individual failure times are known. If the failure times are measured exactly the failures will all occur at separate times, but if they are measured to the nearest month (for example) then there may be several failures at the same time. In this case the probability of failure is estimated by dividing the number of failures at that failure time by the total number of subjects at risk just before the failure time. If losses also occur at this time then, by convention, they are included in the number at risk.

4.5 An example of the Kaplan–Meier method

Table 4.2 shows the time from diagnosis to death from melanoma, or loss to follow-up, for 50 subjects. Times are in complete months so that subjects dying during the first month are recorded as surviving one month, and so on. For two subjects diagnosis took place at death, so the time was recorded as zero.

Note that probabilities of failure are estimated only for times at which failures occurred. The first of these is at time zero; the number at risk is 50, with 2 failures, so the probability of failure at this time point is $2/50 = 0.04$, and the survival probability is $1 - 0.04 = 0.96$. The next time at which a failure occurs is one month; the number at risk is 48, with one failure, so the probability of failure at this time point is $1/48 = 0.0208$ and the probability of surviving is $1-0.0208 = 0.9792$. The next time at which a failure occurs is at 2 months, when there are two failures. The probability of failure is $2/47 = 0.0426$, and the survival probability is $1-0.0426 = 0.9574$. At three months there is one failure and one loss to follow-up. In fact this loss was a death from a cause other than melanoma, but when estimating survival

Table 4.2. Cumulative survival probabilities from the Kaplan–Meier method. Non-melanoma deaths (*) are counted as losses.

				Conditional probability		Cumulative prob.
Month	N	D	L	of death	of survival	of survival
0	50	2		0.0400	0.9600	0.9600
1	48	1		0.0208	0.9792	0.9400
2	47	2		0.0426	0.9574	0.9000
3	45	1	1*	0.0222	0.9778	0.8800
8	43	1		0.0233	0.9767	0.8595
10	42	1		0.0238	0.9762	0.8391
12	41	1	1*	0.0244	0.9756	0.8186
13	39	1		0.0256	0.9744	0.7976
15	38	1		0.0263	0.9737	0.7766
18	37		1*			
19	36	1		0.0278	0.9722	0.7551
21	35		1			
27	34		2			
30	32		1			
33	31	1	1	0.0323	0.9677	0.7307
34	29	1		0.0345	0.9655	0.7055
38	28		1			
40	27		1			
41	26	1		0.0385	0.9615	0.6784
43	25		1			
44	24		1			
46	23		1			
54	22		1			
55	21	1		0.0476	0.9524	0.6461
56	20	1		0.0500	0.9500	0.6138
57	19		2			
60	17		1*			

probabilities from melanoma alone it is counted as a loss to follow-up. (We return to a fuller discussion of this point in Chapter 7.) The number at risk was 45, with one failure, so the probability of failure is $1/45 = 0.022$ and the probability of survival is $1 - 0.022 = 0.9778$, and so on. A plot of the cumulative survival probability against time is shown in Fig.4.7.

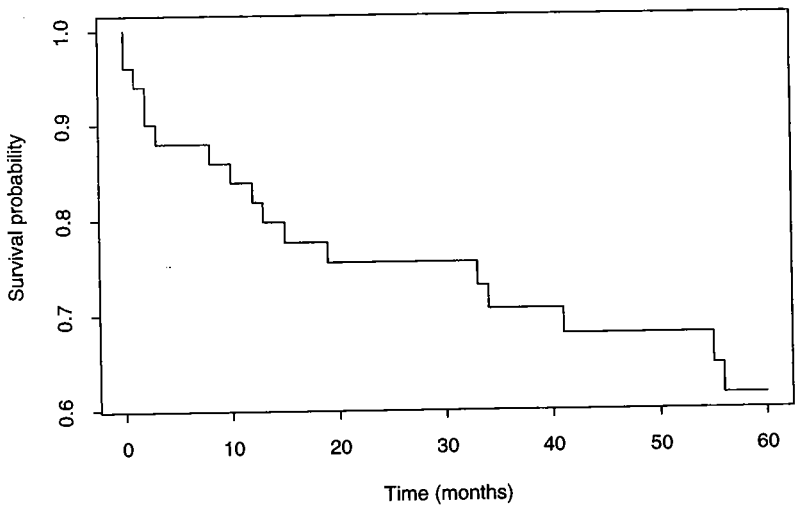


Fig. 4.7. Cumulative survival probability by the Kaplan–Meier method.

Solutions to the exercises

4.1 See Fig.4.8. The probabilities of failure during the first, second and third years are

$$0.05 \quad 0.95 \times 0.09 = 0.0855 \quad 0.95 \times 0.91 \times 0.12 = 0.1037.$$

The probability of surviving three years is

$$0.95 \times 0.91 \times 0.88 = 0.7608.$$

The survival probabilities for the three years are

$$0.95 \quad 0.8645 \quad 0.7608.$$

The probability of failure at some time during the three years is

$$0.05 + 0.0855 + 0.1037 = 0.2392$$

or

$$1 - 0.7608 = 0.2392.$$

4.2 The overall log likelihood is

$$33 \log(\pi) + 232 \log(1 - \pi),$$

which is equivalent to observing 33 failures in 265 subjects. The most likely value of π is, therefore $33/265 = 0.125$.

4.3 See Fig.4.9.

4.4 The five year survival probabilities from Fig.4.5 are 0.78 (Stage I) and 0.51 (Stage II).

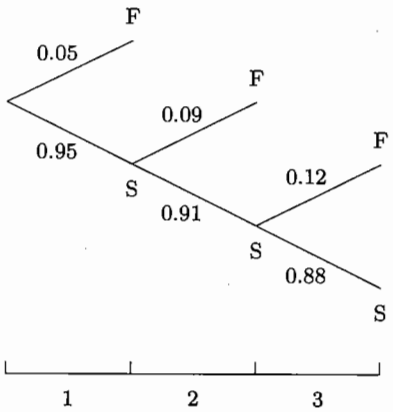


Fig. 4.8. Solution to exercise 4.1.

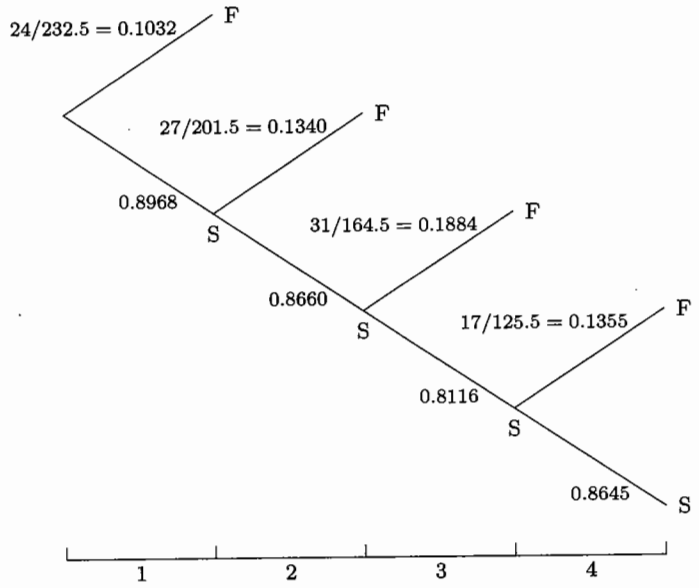


Fig. 4.9. Estimated conditional probabilities for stage II women.

Supplementary Exercise 4.1

Let T be a positive random variable denoting the longevity of a randomly selected product, item, or person (such as an ink cartridge, battery, computer, iPod, or human) or the duration in a given state (e.g., using iOS6, or Windows 8) before transitioning to another state. Denote the associated cumulative distribution function by $F_T(t)$, the survival function by $S_T(t) = 1 - F_T(t)$, the probability density function by $f_T(t)$, the hazard function $f_T(t)/S_T(t)$ by $h(t)$ or $\lambda(t)$, and the expectation $\int_0^\infty t\ f_T(t)\ dt$ by μ_T.

1. Show that
$$\mu_T = \int_0^\infty S_T(t)dt.$$
Mention any textbooks or sources you used to derive this relationship.

   The diagram opposite, which shows the 8221 years lived by 100 persons, can provide some intuition for the proof that relies on changing the order of the integrals. It emphasizes that one can get to the sum of 8221 either by summing the lengths of the horizontal lines – the lifetimes (numbers of years) of the individual persons – or by summing the lengths of the vertical lines – the person years, the collective number of persons alive in each individual year, or as the actuaries say, the 'years lived' in each year. In epidemiology, the latter is by far the more common way. If you get stuck with the general mathematical proof for a continuous T, start with the discrete version, where the logic behind the calculus gymnastics becomes a bit more obvious.

2. Show that
$$S_T(t) = \exp\left[-\int_0^t h_T(u)du\right].$$
   This relationship is also the topic of exercise 4.7 .

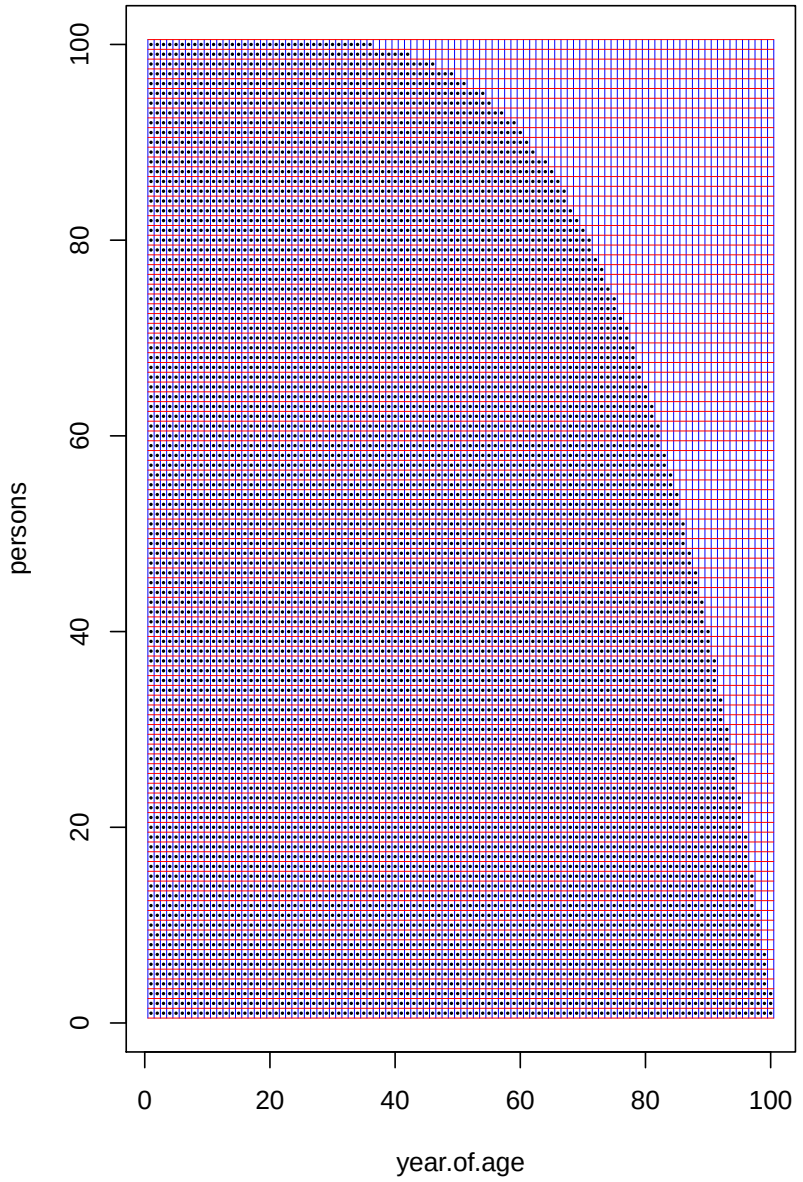


Figure 1: 8221 years lived by 100 persons; mean = 8221years/100persons = 82.21 years/person.

4 Consecutive follow-up intervals

4.1 A sequence of binary models

The lifetable as a sequence of Bernoulli models: Efron (1977) was one of the early authors to point out that the likelihood contribution of a subject, followed for t units of time, is equivalent to the likelihood for a sequence of a large number, $n = t/\Delta$, of Bernoulli trials, with time-dependent probabilities

of failure. For a trial that corresponds to the small interval $(t, t+\Delta)$, the failure probability can be well approximated by $p = h(t)\Delta$, where $h(t)$ is called the hazard function (see later). The sequence ends with the n^{th} trial, at the time of the event of interest or when follow-up was otherwise terminated. In a subsequent article Efron (1988) focused on discretizations of the t-axis and on using logistic regression to fit various smooth-in-t hazard and survival functions in the one-sample situation, where the usual non-parametric alternative is the Kaplan-Meier estimator of survival rate.

The probabilities of surviving one, two, and three years without failing are called the *cumulative survival probabilities* for the cohort: JH continues to argue that the word **'cumulative' is misleading**. The complement of the (unconditional) survival probability is the *cumulative incidence*. When addressing individuals, we call this probability the **Risk**. It is an increasing function of t. Would we call a declining fraction, obtained as a product of more and more fractions, a *cumulative* fraction?

4.2 Estimating the conditional probabilities of failure

The subjects who contribute to the estimation of the conditional probabilities do not have to have been followed from the beginning. One can splice together estimates based on separate samples. This is what is done to create current lifetables. And in any case, when (a subset of) those who "survive" a specific time band are used again in the next band, the estimates are treated as independent of each other – just as if they were from different persons. In current lifetables, they are different persons!

Table 17.1 in p. 570 of the Survival Analysis chapter (17) of the 4th edition of Statistical Methods in Medical Research by Armitage, Berry & Matthews (see opposite) nicely illustrates the difference between 'current' (a.k.a. 'period') and 'cohort' lifetables.

The entire 'current' lifetable is calculated, as a product of conditional probabilities, using the *observed* age-specific mortality rates in England and Wales in 1930-1932. In this sense it is fictitious, since those who computed the table in the 1930's didn't know for sure that the world would even exist in 2010, when those remaining from the fictional 1000 who started out at age 0 would reach their 80th birthday. And even if they did, they could not have anticipated exactly what force of mortality these 80-year olds would face in 2010, even though they might have foreseen that mortality rates would improve over time. The force of mortality these 80-year olds would face in 2010 is a good deal lower than the force of mortality the 80-year olds actually faces in 1930-32. For example, the death rate in the male 75-79 age category in Denmark was 9.4/100MY in 1930-34 and 4.2/100MY in 2000-04.

570 Survival analysis

Table 17.1 Current and cohort abridged life-tables for men in England and Wales born around 1931.

	Current life-tables 1930–32			Cohort life-table, 1931 cohort
Age (years) x	Probability of death between age x and $x+1$ q_x	Life-table survivors l_x	Expectation of life $\mathring{e}_x$	Life-table survivors l_x
0	0·0719	1000	58·7	1000
1	0·0153	928·1	62·2	927·8
5	0·0034	900·7	60·1	903·6
10	0·0015	890·2	55·8	894·8
20	0·0032	872·4	46·8	884·2
30	0·0034	844·2	38·2	874·1
40	0·0056	809·4	29·6	861·8
50	0·0113	747·9	21·6	829·7
60	0·0242	636·2	14·4	—
70	0·0604	433·6	8·6	—
80	0·1450	162·0	4·7	—

> "The cohort life-table describes the actual survival experience of a group, a 'cohort' of individuals born at about the same time. Those born in 1900, for instance, are subject during their first year to the mortality under 1 year of age prevailing in 1900-1; if they survive to 10 years of age they are subject to the mortality at that age in 1910-11; and so on. Cohort life-tables summarize the mortality at different ages at the times when the cohort would have been at these ages. The right-hand side of Table 17.1 summarizes the l_x column from the cohort life-table for men in England and Wales born in the 5 years centred around 1931. As would be expected. the values of l_1 in the two life-tables are very similar, being dependent on infant mortality in about the same calendar years At higher ages the values of l are greater for the cohort table because this is based on mortality rates at the higher ages which were experienced since 1932."

For a further illustration of the difference between 'current' and 'cohort' life tables, see the Bridge of Life applets (`https://jhanley.biostat.mcgill.ca/BridgeOfLife/`). In particular, see the contrast between France, 1895 (current) and France, 1895-2004 (cohort).

*This exercise makes it clear that, in the analysis of such studies, the **basic atom of data** is not the subject, but the observation of **one subject through one time band**.* [last para of section 4.2]

This is a very important statement, and this 'outlook' or 'attitude' is key to a full understanding of rates, and or person-time. It says that one's 'timeline' is **divisible**. Think of the experience as an infinite sequence of Bernoulli trials that is terminated by the event, or when observation is terminated (i.e., before the event could occur).

It also allows the experience to be further sub-divided into 'exposed' person time bands and 'unexposed' person time bands: c.f. of the 'clicks' of time a driver spends on the cell-phone and off-the-cell-phone.

In the example, the event of interest is a **one-time event**, and so, unlike the cat with nine lives, once the event occurs, it terminates the observation: one is no longer 'at risk.' But one can also think of events, such as repeated events such as accidents, or sickness episodes, experienced by the same person.

4.3 A cohort life table

These [survival] plots are useful for studying whether the probability of failure is changing with follow-up time, and for calculating survival probabilities for different periods of time. In fact, it is not that easy to check if the probability of failure is changing from survival curves. The probability of failure the authors write of is a conditional, i.e. time-specific, probability, and so the hazard function, which uses as a denominator the numbers of persons at risk at that time, makes it easier to monitor this probability.

4.4 The use of exact times of failure and censoring

"[...] *choosing the bands so short that each failure occupies a band by itself.*" This is the same assumption that allows us to derive the Poisson distribution as a limiting case of the Binomial distribution, and the link between the Poisson distribution and the exponential distribution of inter-event times.

"*This method of estimating the cumulative survival probabilities is called the Kaplan-Meier method*" It is also called the **product-limit method**, since it is derived by slicing time into smaller and smaller bands, and not having to be materially concerned about where within the band an observation becomes censored. In the JUPITER trial example JH is using in the EPIB-634 course, the follow-up ranges from just over a year to almost 5 years, or approximately 400 to 1600 days. The 200+ events in the placebo arm, and the 100+ in the treatment arm, are distributed over these 1600 days. If we use one day as the width of each band, and estimate $S(1000)$, the 1000-day "event-free survival" then this estimate is a product of 1000 estimated conditional probabilities, many of them estimated at unity. *So the changes in the product take place only at the days in which there were events.* See also the COMPARE trial

> *In this case, the probability of failure is estimated by dividing the number of failures at that failure time by the total number of subjects at risk just before the failure time.* The persons at risk just before the event on a particular day (including the person(s) who did suffer the event that day) are called the ***riskset***. They are the ***candidates*** for the event.

"If losses also occur at this time then, by convention, they are included in the number at risk." middle of p. 35

This sometimes causes confusion for end-users and even some epidemiology teachers, and JH has often had epidemiology students ask him why it is. His answer is that if time were truly 'continuous', it is unlikely that we would have losses to follow up and events at the same 'exact' moment. The issue arises when we deal with discrete times, either because they have been rounded or binned, or are naturally discrete (e.g., how many years or terms of fees one pays while in the graduate program, or for how many cycles a couple must 'try' before they conceive a child. In such cases, it is usually quite clear: a $t = 5$ means success on cycle 5, or a PhD by the end of year 5, while a $t = 5+$ means the milestone or goal was not achieved in 5 time units. Clearly, those persons for whom a $t = 5+$ is recorded were (unsuccessful) 'candidates' in the 5th trial. Notice here that becoming pregnant or graduating is not a (statistical) failure: it is merely a transition from one state (not pregnant, still pursuing a PhD) to another (pregnant, PhD).

So it is not just by convention, but rather by logic and common sense, that they are included in the risk set.

Supplementary Exercise 4.2

Consider again the tumbler longevity data that we saw in an earlier exercise. The smallest unit of time (the 'granularity') is 1 week. Even though some observations ($< 10\%$) are right-censored, Table 1 in the paper lists the data in a form that allows direct calculation of an empirical complement-of-the-cdf or 'K-M' estimate; You might wish to call your method the **'coarse-'** rather than **'exact-'** product-limits curve.

1. Graphically compare the $\widehat{S(t)}$ obtained with this (non-parametric) 'K-M' estimator of 'the survival' function with the results obtained with the (parametric) gamma model fitted by the authors. (Of course, if in your

K-M curve, you take failures to have occurred at the very end of each week, just before the person came around to check on them, and plot the resulting step function that drops at the end of each week, your 2 curves are bound to disagree somewhat *within* each week)

2. Compare the mean longevity estimated by calculating the area under this non-parametric (K-M) survival curve (see exercise 4.1) with the fitted mean obtained from the values of the 2 fitted parameters of the author's model. Give reasons why they differ.

3. What if the inspection times were daily (hourly) rather than weekly, and the failure and censoring times correspondingly finer? (Approximately) how many jumps (and, thus, how many consequential multiplications) would there be in the **'exact-'** product-limits curve if the inspections were (a) daily (b) hourly (c) even more exact?

4.4.1 $\widehat{S(t)}_{KM}$ is a Maximum-Likelihood estimator of $S(t)$

As is rigorously justified in their 1958 paper, the Kaplan-Meier estimator is a non-parametric ML estimator within the class of all possible $S(t)$ functions.

Supplementary Exercise 4.3

Take a small survival dataset with just 3 observations, 1 censored and 2 not, such as the 3 values 5, 7+ and 10. Show that

$\widehat{S(t)}_{KM}$	Interval	Point (t)	Prob. Mass at Point
1	$t < 5$		
		$t = 5$	1/3
2/3	$5 \le t < 10$		
		$t = 10$	2/3
0	$t \ge 10$.		

maximizes the Likelihood, ie the probability of the observed data as a function of $S(t)$, i.e., that no other $\widehat{S(t)}$ can yield a larger likelihood.

4.4.2 $\widehat{S(t)}_{KM}$ as a 'self-consistent' and as a 'Distribute mass to the right' estimator of $S(t)$ – Efron, Berkeley Symposium, 1967

See the full article. The K-M estimator, based on n observations $T_1, \ldots, T_n$, some censored, some not, can also be seen as obeying the **self-consistent** estimating equation:

$$S(t) = \frac{1}{n}\left\{ \sum_{all} I[T_i > t] + \sum_{censored\ <\ t} \frac{S(t)}{S(T_i)} \right\}$$

Observations known to exceed t [even if censored after t] are counted as survivors (1's) while observations for which we don't know if they will exceed t are counted as fractions or probabilities: those which are already close to reaching t are given higher chances of eventually exceeding it, those which are further to the left of t are given lower chances of doing so: the chance of eventually exceeding t, given that one has already reached a value $T < t$, is $S(t)/S(T)$.

As explained in the same 1967 article, the K-M estimator can also be seen as a **distribute to the right** procedure: Initially, each of the n observations is given a mass of $1/n$. Then, the mass given to the leftmost censored observation is redistributed (equally) to all observations to the right of it, and that leftmost observation is removed. The process is repeated until all censored observations are removed, and all of their mass has been redistributed. [1]. The procedure will remind some of the EM algorithm, déjà vu.

Supplementary Exercise 4.4

Take a simple survival dataset with just 5 observations, 2 censored and 3 not, such as the 5 values 2, 5+, 6, 7+ and 9. Derive the K-M estimate of $S(t)$. Illustrate the 'self-consistency' of the KM estimator, and that the 'distribution to the right' procedure produces the KM estimate. You might wish to consult the excellent teaching article 'Kaplan-Meier Theatre' by Thomas Gerds, available here.

Supplementary Exercise 4.5

The self-consistent property can also be used with more complicated censoring, such as interval censoring and – as the most extreme case – 'current status' data (e.g., the avalanche dataset) where each observation is either left-censored (dead when extracted) or right-censored (alive when extracted)

Exercise: Consider a dataset with 10 observations: the *true* values have no time element, but are (possibly repeated) prime numbers between 1 and 29

[1] `https://jhanley.biostat.mcgill.ca/bios601/SurvivalAnalysis/Efron1967.pdf`

inclusive. 6 are left-censored (<10, <16, <18, <21, <26, <28) and 4 are right-censored (>6, >10, >11, and >24.

Analytically, and separately by repeated (iterative) use of the 'self-consistency' principle, arrive at an estimate of $S(t)$.

Hint: You may find the diagram produced by the supplied R code (see website) helpful to visualize the data-intervals.

Start by choosing the support points (here integers) over which the total of probability mass of 1 will be distributed. Try to have these integer values [points of 'support'] be as helpful as possible – include them in (and thus make them contribute to the likelihood of) as many of the data-intervals as possible. In this example, the minimal set of support points has size 3 (note that the 3 points are not unique).

Analytically: write down the likelihood as a function of the magnitudes, θ_1, θ_2, and $\theta_3 = (1 - [\theta_1 + \theta_2])$ of these 'parameters.' Then maximize this with respect to θ_1 and θ_2, say.

Iteratively: Start by strategically selecting 3 probability masses $\{\theta_1^{[0]}, \theta_2^{[0]}, \theta_3^{[0]}\}$ to distribute over the 3 selected support points. This distribution gives you an initial estimate, $S_0(t)$, of the $S(t)$ function. (Out of interest, calculate the Likelihood associated with this $S(t)$).

Then use this $S_0(t)$ as the $S(t)$ in the right hand side of the equation at the beginning of section 4.4.2 to obtain a new estimate, $S_1(t)$ of the $S(t)$ function. (again, out of interest, calculate the Likelihood associated with this new $S(t)$)

Repeat until the estimate of the $S(t)$ function (and the Likelihood) no longer changes.

Supplementary Exercise 4.6

Use the supplied `R` code (or 'roll your own' code) to obtain a NPMLE of the $S(t)$ function in the case of the avalanche data.

Look on the web for software that can do this, and tell us what you were able to find, and how flexible and user-friendly it appears to be.

4.4.3 The Nelson-Aalen estimator of $S(t)$

This is also covered in Chapter 5.6 of Clayton & Hills.

Just as with K-M, divide the entire interval $[0, t]$ into J *narrow* event-containing sub-intervals; ignore the 'non-event-containing' sub-intervals.

Sub-interval j is defined by distinct event-time t_j, with n_j at risk just before the event(s) [death(s)] in that interval. (there can be more than 1 event at the same t_j, particularly if time is measured coarsely).

The (step-)function $n(t)$ is the number at risk at each time point in $(0, t)$. **'Riskset'$_j$ = the n_j 'candidates' for the event(s) at t_j.**

Suppose s_j survive event-containing sub-interval j, and that the remaining $d_j = n_j - s_j$ do not [the letter d is used here because in many applications, the 'transition' ('event') is from the initial state of 'alive' to the destination state of '$d$ead', but transitions may be desirable or undesirable.]

The Nelson-Aalen Estimator uses the same general formula that links the $S(t)$ and $ID(t)$ or $\lambda(t)$ functions:

$$\widehat{S_{NA}(t)} = \exp\left\{ -\int_0^t ID(u)du \right\} = \exp\left\{ -\int_0^t \lambda(u)du \right\} = \exp\left\{ -\sum \frac{d_j}{n_j} \right\}$$

Think of a fitted ID function $ID(t)$ with $\widehat{ID(t)} = 0$ in the non-event-containing sub-intervals of $(0, t)$ and $\widehat{ID(t)} = d/PT = d/(n \times \delta t)$ in each event-containing interval of width δt; thus $\widehat{ID(t)} = d_j/(n_j \times \delta t)$ in event-containing interval j.

Supplementary Exercise 4.7

Read the manuscript 'From incidence function to risk' which JH submitted to the American Journal of Epidemiology on 2013.05.06. It can be found here.

This manuscript is a consolidation of drafts of two earlier separate teaching articles, which (if your are interested *and* have the time) can be found under "Farr 'On Prognosis'; Vandenbrouche/Morabia on 'risks and rates' " tab in the Website for course EPIB609 `https://jhanley.biostat.mcgill.ca/c609/material/index.html#RisksRates`. Part I, which will be more relevant to C&H Chapter 5, addresses incidence density, force of mortality, and hazard functions, while part II (the basis for the AJE submission), more relevant to C&H Chapter 4, attempts to explain the link between the $S(t)$ and $ID(t)$ or $\lambda(t)$ functions.

Of course, for biostatisticians who are quite comfortable with integral and differential calculus, the fact that $S(t) = \exp\{ -\int_0^t \lambda(u)du \}$ can be established directly by solving the differential equation that defines the hazard function, $\lambda(t)$, or if you prefer the letter h, between $S(t)$ and $h(t)$. From the basic definition of the hazard as a conditional failure rate, we have that $h(t) = f(t)/S(t)$, where f is the pdf, and S the complement of the cdf. Since $f(t) = -S'(t)$, we can write $h(t) = f(t)/S(t) = -S'(t)/S(t)$. Solving this differential equation in $S(t)$ leads immediately to the link.

.

Even though he did not mention it in class when we were covering the link

between the Poisson the exponential waiting-time distributions, JH has since used the idea of 'generations' to construct a 'Poisson Generations' graph. You can find it under the Lecture notes in the resources for statistical models [intensity].

.

For this exercise, you are asked to summarize in a few sentences of your own the purpose and content of the 2013.05.06 version 'From incidence function to risk', and to say whether you think the article would be helpful even for people who are very comfortable with integral and differential calculus. Also, if you see any places where you think the exposition/writing can be improved, do not be shy in saying so. And in particular, JH would welcome any comments on how the Nelson-Aalen estimator is described and whether the article demystifies it in any way.

Supplementary Exercise 4.8

1. Using the $\widehat{ID(t)}$ function described in section 4.4.3, evaluate the integral of $\int_0^t \widehat{ID(u)}du$ and use it to obtain the Nelson-Aalen estimator of $S(t)$.
2. Derive the conditions under which the K-M estimator $\prod \frac{s_j}{n_j} = \prod\{1-\frac{d_j}{n_j}\}$ gives a result that is very close to that of the Nelson-Aalen estimator.
3. Assuming $d_j \sim Poisson(\mathrm{E}[d_j])$, derive an expression for $Var[\widehat{S(t)_{NA}}]$.[2]
4. Report on your small survey of the web (or textbooks) as to how many use this Poisson-based version for the components of $Var[\widehat{S(t)_{NA}}]$ and to how many use the binomial-based version for them. Which makes more sense to you?

Supplementary Exercise 4.9

Refer to the data, contained in Figure 4 in this unpublished manuscript. It addresses the frequency of IUD discontinuation because of bleeding.

1. Just from the figure, determine the sizes (and time-locations) of the 9 risk sets.
2. Using these, and by hand, reproduce the K-M and N-A $\widehat{S(t)}$ values (and their SE's) at the first 3 'jumps'.
3. Suggest a label for the '$S(t)$' axis.
4. Give reasons for presenting a plot of the function $1-\hat{S}(t)$, rather than the function $\hat{S}(t)$.
5. Suggest a label for the '$1-\hat{S}(t)$' axis.
6. Use the `surfit` function in the `survival` package in `R` to derive the K-M and N-A estimates. The data (and incomplete `R` code) are in the 'Resources' link for Chapter 04 of C&H. (Duration (weeks) before discontinuation(denoted by a '1') of IUD (data from Collett)). The details are in the links to `survfit.formula` etc. [3]

[2] Hints: First, work out the variance of $\log\widehat{S(t)_{NA}}$ first, and then the variance of its antilog. Use the relationship $\mathrm{var}[d_j] = \mathrm{E}[d_j]$, and use d_j as a plug-in estimator for $\mathrm{E}[d_j]$.

[3] Oddly enough, according to `http://www.ats.ucla.edu/stat/r/examples/asa/asa_ch2_r.htm` "The easiest way to get Nelson-Aalen estimator is via cox regression using coxph function." `survfit(coxph(Surv(time,censor) ~ 1), type="aalen")`

4.5 Examples of the Kaplan-Meier method

Example 1 Cf. JUPITER data on the website for course EPI634

The `R` code calls the "canned" routines, but also derives the K-M-based cumulative incidence curves 'from scratch.'

Example 2 Figure 2 below is from the article: "Male circumcision for HIV prevention in young men in Kisumu, **Kenya**: a randomised controlled trial" (Lancet 2007; 369: 643-656). The full article is here. There is also find a companion article for a similar randomized trial, with similar estimates of benefit, carried out in Uganda,

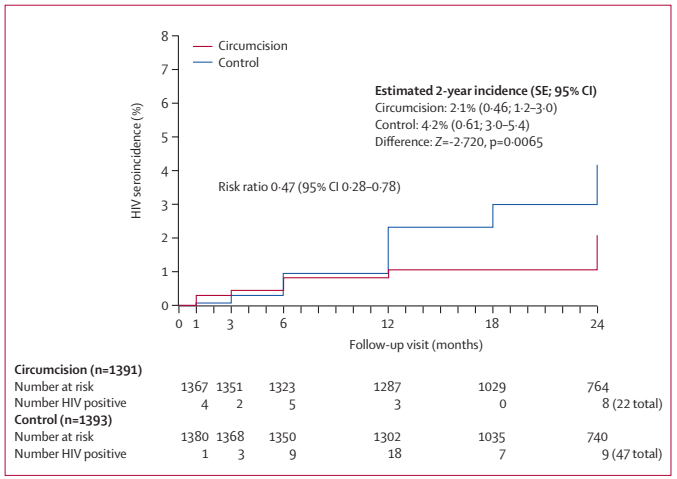


Figure 2: **Cumulative HIV seroincidence across follow-up visits by treatment**
Time to HIV-positive status is taken as the first visit when a positive HIV test result is noted. Time is credited as the follow-up visit month. Participants without HIV-positive status are censored at the last regular follow-up visit completed where HIV testing was done, credited specifically as months 1, 3, 6, 12, 18, and 24.

.

Supplementary Exercise 4.10

.

Replicate the statistics reported in the insert beginning with the text "Estimated 2-year incidence" in the top right portion of the above Figure 2.

	Intervention group	Control group	Incidence rate ratio (95% CI)	p value
0–6 months follow-up interval				
Number of participants	2263	2319		
Incident events	14	19		
Person-years	1172·1	1206·7		
Incidence per 100 person-years	1·19	1·58	0·76 (0·35–1·60)	0·439
6–12 months follow-up interval				
Number of participants	2235	2229		
Incident events	5	14		
Person-years	1190·7	1176·3		
Incidence per 100 person-years	0·42	1·19	0·35 (0·10–1·04)	0·0389
12–24 months follow-up interval				
Number of participants	964	980		
Incident events	3	12		
Person-years	989·7	1008·7		
Incidence per 100 person-years	0·30	1·19	0·25 (0·05–0·94)	0·0233
Total 0–24 months follow-up				
Cumulative number of participants	2387	2430		
Cumulative incident events	22	45		
Cumulative person-years	3352·4	3391·8		
Cumulative incidence per 100 person-years	0·66	1·33	0·49 (0·28–0·84)	0·0057

Table 3: **HIV incidence by study group and follow-up interval, and cumulative HIV incidence over 2 years**

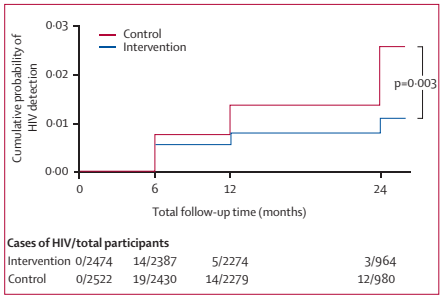


Figure 2: **Kaplan-Meier cumulative probabilities of HIV detection by study group**
Actual visits grouped by the three scheduled visits at 6 months, 12 months, and 24 months after enrolment. The cumulative probabilities of HIV infection were 1·1% in the intervention group and 2·6% in the control group over 24 months.

Example 3 The items above are from "Male circumcision for HIV prevention in men in Rakai, **Uganda**: a randomised trial," Lancet 2007; 369: 657-666.

.

Supplementary Exercise 4.11

.

Comment on the appropriateness of (i) the term "*Cumulative incidence per 100 person-years*" in the last row of Table 3 (ii) using a single incidence (hazard) rate ratio of 0.49 for the full 2 years, and in the abstract, reporting the estimated efficacy of intervention as 51%.

Supplementary Exercise 4.12

Refer to the 2015 article "Feasibility and effectiveness of oral cholera vaccine in an urban endemic setting in Bangladesh: a cluster randomised open-label trial" published in The Lancet, and to this R code

1. In the 'Statistical analysis' section, the authors state that

   > We calculated sample size by methods described elsewhere.[*Donner&KlarText*] We calculated the intra-cluster correlation for cholera hospital admissions for 2008, and 2009. We assumed 65% efficacy and 65% coverage, yielding 42% overall protective efficacy, with a one-sided test ($\alpha = 0.05$), 80% power, incidence of 1.6 cases per 1000 people per year, 25% yearly attrition, and 2 years of post-vaccination surveillance. On the basis of these assumptions, we calculated that we would need 236,340 participants (78,780 in each group).

   (a) What (assumed constant) attrition rate [expressed as an instantaneous rate of say x losses per 100 participant-days] would generate an annual attrition risk of 25% (so that only 3/4 of those randomized remain under followup ['at risk'] at the end of year 1, and only 9/16 at the end of year 2)?

   (b) Assume 50,000 persons were to be randomized to the control arm. Using the attrition rate you just calculated, compute and graph the expected number under surveillance for each of the first 730 days of follow-up. Add up these daily numbers to get the expected total number of person-days (PD) of follow-up.

   (c) For the moment, ignore the variance (reciprocal of sample size) inflation caused by the cluster randomization design, (i.e., natural cluster to variation in attack rates) and by the fact that cholera can also easily spread between persons in the same cluster. Assume instead that the individual attacks are governed by a single homogeneous Poisson process, with $\lambda_0 = 1.6/1000$ PersonYears (PY) in the 'control arm' area.

   Convert this attack rate to an attack rate per person-day or per 100,000 person-days. {For interest, is it higher or lower that the observed rate given in Table 2?}

   Let Y_0 denote the number of attacks in the control arm. Using the rate you just calculated, and the PD from above, calculate E[Y_0], and (under the no extra-variance assumption) Var[Y_0].

   (d) Assume, for the purposes of hypothesis testing and setting a 'positivity' cutoff for a statistical test of the null, that the attack rate (λ_1) for persons in the 'vaccination only' area is also 1.6/1000PY, and that the 50,000 persons randomized to this arm are subject to the same attrition rate. Let Y_1 denote the number of attacks in this arm, and let $d = Y_1 - Y_0$.[4] Under this null assumption, calculate $\sigma_{d|H_0} = Var[d|H_0]^{1/2)}$, and compute

   $$d_{crit} = E[d|H_0] - 1.96 \times \sigma_{d|H_0} = 0 - 1.96 \times \sigma_{d|H_0}.$$

   (Note that this implies a 2-sided test with $\alpha = 0.05$; it appears that the authors used a 1-sided test, so they would have used 1.645 SD's as their criterion for test positivity). Sketch the distribution of $d|H_0$, leaving some space to the left of, and below, the distribution so as to be able to overlay the non-null version. (Given the large expected numbers, it is safe to use a Normal approximations to the exact distributions of $d|H_0$ and $d|H_1$.)

   (e) Under the authors' assumptions, what is the (non-null, H_1) value of λ_1, of d, and of $\sigma_{d|H_1}$? Sketch this distribution of $d|H_1$, to the left of the null distribution, and upside down[5] so it is easier to distinguish the various tail areas. Then use a normal approximation to the distribution of $d|H_1$ to calculate what percentage of it lies to the left of d_{crit}. This percentage is called the *power* of this size study, i.e., the probability that – assuming the λ_1 and λ_0 are as specified – the study will yield a 'positive' (i.e., statistically significant) result.[6]

2. In the 'Results' section, the authors address a measure of the 2-year protection afforded by the vaccine. They had two ways of obtaining a crude measure:
   (i) as $100 \times (1 - RiskRatio)$, where the $RiskRatio$ is estimated as the ratio of the 2-year (730 day) risks (i.e. approximately $\widehat{R_1(2y)} = 1 - 0.9989 = 0.0011$ and $\widehat{R_0(2y)} = 1 - 0.9981 = 0.0019$) obtained from the the two Kaplan-Meier curves in Figure 3). This gives a crude estimate of $100 \times (1 - 0.0011/0.0019) = 42\%$ protection.

[4] If the 2 amounts of experience were different, we would need to consider the difference in the *rates*, rather than in the numerators. This 'close to 50:50' randomization makes the planning a bit easier.

[5] See diagram in section 4.3.2 (p14) of JH's Notes on inference for a mean. It is a simpler (one-sample) context, and the alternative is on the positive side of the null, but it gives you the idea. For more on sample size calculations for a comparison of 2 means, see the Notes on comparison of 2 means in the Resources.

[6] It does **not** mean, as some investigators sloppily write, that the study has this power to *detect* a rate ratio reduction of 42%.

(ii) as $100 \times (1 - RateRatio)$, or $100 \times (1 - HazardRatio)$, where the $RateRatio$ or $HazardRatio$ is estimated as the ratio of the attack rates calculated over the 730 days (i.e. approximately $\widehat{\lambda_1} = 65/41,809,947PD = 0.1555$ attacks per 100000PD and $\widehat{\lambda_0} = 106/39,327,744PD = 0.2695$ attacks per 100000PD obtained from the data in the top row of Table 2). This also gives a crude estimate of $100 \times (1 - 0.1555/0.2695) = 42\%$ protection.

(a) Assuming no extra-Poisson variation, we have enough information to directly calculate a CI for the second version. We start by calculating a CI for the $RateRatio$ or $HazardRatio$, and then compute $100 \times (1 - CI)$. But instead of working in the ratio scale, we work in the $\log[Ratio]$ scale, so that $Var\{\log[\widehat{\lambda_1}/\widehat{\lambda_0}]\} = Var\{\log[\widehat{\lambda_1}]\} + Var\{\log[\widehat{\lambda_0}]\}$.

Use your results from exercise 0.1 of the 'models for intensity rates' material to work out the variance (and thus a CI) for the log of the ratio, and from it, a CI for the ratio itself. Then, convert this (slightly asymmetric) CI for the ratio into a (similarly asymmetric) CI to accompany the point estimate of the percent protection. Can you think of reasons why their CI is slightly wider?

(b) With a few approximations and interpolations, and again assuming no extra-binomial variation, we have enough information to directly calculate a Greenwood-based CI for the first version.

$$\widehat{RiskRatio} = \widehat{R_1(2y)}/\widehat{R_0(2y)}$$

$$\text{Var}\{\log \widehat{RiskRatio}\} = \text{Var}\{\log[\widehat{R_1}]\} + \text{Var}\{\log[\widehat{R_0}]\}$$

Writing $S = 1 - R$, noting that $\text{Var}[\hat{R}] = \text{Var}[1 - \hat{R}] = \text{Var}[\hat{S}]$, and using the delta method, we can write each of the two Var{log} components as

$$(1/\hat{R})^2 \times \text{Var}[\hat{R}] = (1/\hat{R})^2 \times \text{Var}[\hat{S}] = (1/\hat{R})^2 \times \hat{S}^2 \times \sum\{n_i^{-2}\}$$

where the sum is over the risksets.[7] In this application, each riskset, i.e., n_i, is very large; if the attacks occur on separate days, so that each d_i is 1, then each $d_i/[n_i(n_i - d_i)]$ term in the sum in the Greenwood formula can be approximated by $1/n_i^2$. So, all we are missing for the two components are the 65 different n's, i.e. the numbers at risk in the vaccination arm when each of the 65 attacks occurred, and the 106 numbers at risk in the control arm when each of the 106 attacks occurred. Figure 3 indicates that the events are distributed across the 730 days, but because there is more person time nearer to day 1 than day 730 (see your first set of calculations), the 65 events are too. and so the n's at these times tend to be somewhat bigger than the average n.

Generate a best guess for the 65 n's at risk, and from them calculate the first variance component for $\text{Var}\{\log \widehat{RiskRatio}\}$. Do the same for the other arm (with 106 attacks), and add the two variance components to get the variance of the $\log RiskRatio$. From this, calculate a CI for $\log RiskRatio$ and, from it, a CI for $RiskRatio$, and, from it, a CI for the $Percent\ Protection$.

3. Refer again to the 'Statistical analysis' section, where the authors address the intra-cluster correlation for hospital admissions.

Assume, for simplicity, that all clusters have the same (average) cluster size of $n = 9,001$, so that the variance inflation factor[8] is $VIF = 1 + (n-1) \times ICC = 1 + 9000 \times ICC$.

Using the same steps as in the power calculation above, and assuming for now that ICC=0, work out what sample size per arm would be required[9] for the type II error of 20% (80% power) if one uses a test of size $\alpha = 0.05$ (1-sided).

Compare this with the requirement calculated by the authors, and deduce the value of ICC they must have used.

[7] Each riskset is the candidates for the attack in question, and we assume for simplicity that each attack occurred on a different day, so there are no 'ties'.

[8] For a derivation of this relationship, refer to section 3 'Power / Precision / Sample Size: Correlated responses; cluster samples' of the Notes on Comparisons of 2 Means: - models / (frequentist) inference / planning, which can be found under the 2 MEANS section (bottom of page) of the Resources: Models/ Inference / Planning [mean/quantile].

[9] You may also find the 'Rate ratios' section in the article, 'Sample Size, Precision and Power Calculations: A Unified Approach' by Hanley and Moodie in J Biomet Biostat 2:124 in 2011 (link) to be of help. It calculates the requirements in terms of *numbers of events*, but it is easy to work back from numbers of events to the numbers of person-days required to generate this many events.

Supplementary Exercise 4.13 Recovering the data behind a Kaplan-Meier curve

The 'Marriage-free survival' figure below is taken from the article 'Marriage risk of cancer research fellows' in the 2011 Christmas Edition of The Lancet.

The authors[10] begin by telling us

> Research fellows aiming to obtain a PhD or MD/PhD degree face many hazards at work, including exposure to toxic substances and harassment by reviewers of their papers. However, few data exist on the sociocultural risk factors encountered at work – eg, their risk of marriage.
>
> Therefore, between 1993 and 2008, we entered all our 13 research fellows (12 men, one woman; median age 29 years [range 26-32] at study entry; median length of stay in the laboratory 36 months [18-42]) into this prospective, observational, happily-matched-pair cohort analysis. The primary study endpoint was the date of marriage of a research fellow recorded by the respective Swiss Departments of Administrative Affairs and Marital Matters. We took great care not to influence the partner choice of our fellows. Quality-of- life assessment was deemed to be superfluous given the happy faces of study participants recorded when they reached the trial endpoint.

and then report

> **11 of 13 participants** (85%) got married by the 17-year cutoff (figure) – ie, when this report was prepared by one of us (MFF) during a Swiss railway journey to attend a study participant's wedding. Two research fellows are still at risk, but we are confident (unpublished data) that they, too, will eventually reach the endpoint. No toxic effects were recorded, which is remarkable for an oncology trial.

They comment that

> Young academics embarking on a research fellowship in experimental oncology run an excessive risk of ending up in marriage before or shortly after having obtained their MD or PhD degree. The Swiss Federal Office for Statistics indicates that, in our population, the overall risk of living in a married state is 44.5%, and the age-adjusted risk (with respect to our fellow population) is only 38.6%.[1] We therefore felt that, by Kitchen's criteria on statistical proof of the bloody obvious,[2] our results were so clear-cut as to obviate the need for any significance tests. Clandestine data collection ensured that participants were not influenced by us when undertaking steps to shape their social structures and emotional networks. Any cohort study design can be criticised for bias, but a randomised intervention trial would have led to protocol violations whenever love?s labour?s lost, and might have ruined the fun of finding a suitable partner.
>
> Our landmark findings indicate that research fellows must be fully informed of this potential hazard when making up their mind as to whether or not to embark on an academic degree in experimental cancer research.

References

1 Swiss Federal Office of Statistics. Characteristics of the Swiss population by marital status.

`http://www.bfs.admin.ch/bfs/portal/de/index/themen/01/02/blank/key/zivilstand.html` (accessed Sept 9, 2011).

2 Kitchen I. Statistics and pharmacology: the bloody obvious test. Trends Pharmacol Sci 1987; 8: 252?53.

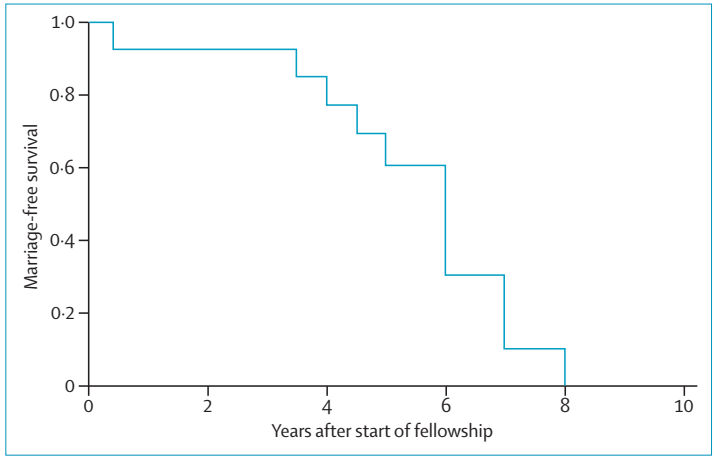


Figure: **Marriage-free survival**
Individuals with a bachelor status were censored at the time of analysis.

Questions for `bios601` students: [11]

1. Determine the times of the 11 marriages.
2. Determine (as best you can) how long the two fellows had (still been) at risk when the report was prepared.

[10] Martin F Fey, Andreas Tobler; martin.fey@insel.ch. Department of Medical Oncology (MFF) and Department of Haematology (AT), Inselspital, University Hospital of Berne, 3010 Bern, Switzerland. Both authors claim equal rights on first and senior authorships. We declare that we have no conflicts of interest.

[11] The article '*Recovering the raw data behind a non-parametric survival curve*,' is available from `https://www.ncbi.nlm.nih.gov/pubmed/25551437` or from here.

Supplementary Exercise 4.14 Can you settle an argument?

The data behind Figures 1 and 2, [12]

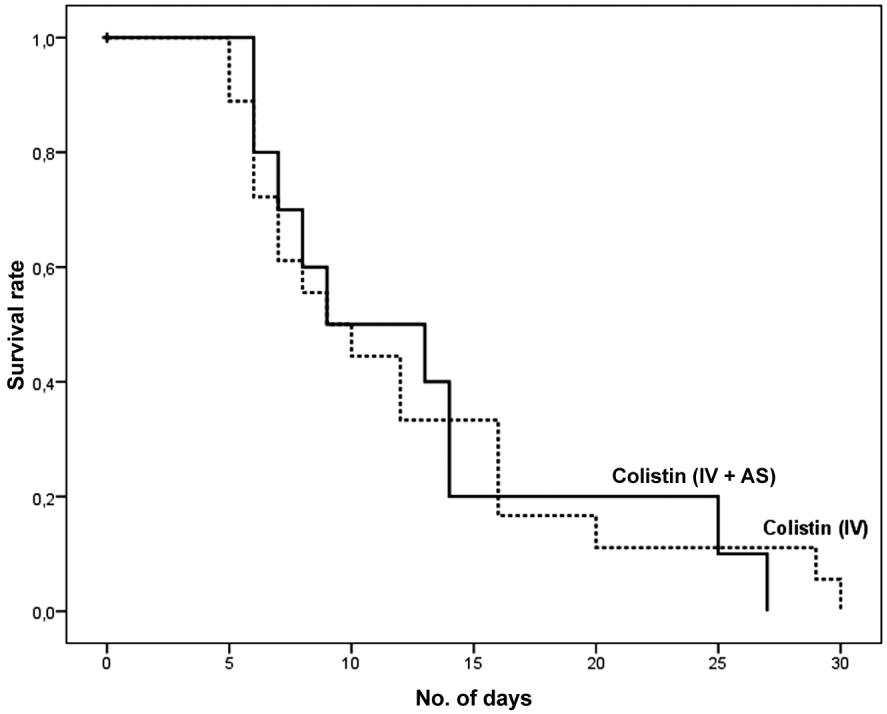


Figure 1. All-cause mortality in the 2 treatment groups. AS, aerosolized; IV, intravenous.

[12] from the article 'Aerosolized plus Intravenous Colistin versus Intravenous Colistin Alone for the Treatment of Ventilator-Associated Pneumonia: A Matched Case-Control Study' by Kofteridis et al. in Clinical Infectious Diseases 2010;51(11):1238-1244 (link)

Objectives. The incidence of ventilator-associated pneumonia (VAP) due to multidrug-resistant (MDR) organisms is increasing. Intravenous (IV) colistin or aerosolized (AS) plus IV colistin have been recently used to treat these life-threatening infections. The purpose of this study was to compare the efficacy and safety of AS plus IV colistin versus IV colistin alone for patients with MDR VAP due to gram-negative bacteria.

Methods. A retrospective matched case-control study was performed at the Intensive Care Unit of the University Hospital of Heraklion, Greece, from January 2005 through December 2008. Forty-three patients with VAP due gram-negative MDR pathogens received AS plus IV colistin and were matched on the basis of age and Acute Physiology and Chronic Health Evaluation II score with 43 control patients who had received IV colistin alone.

Results. Demographic characteristics, clinical status, and gram-negative isolated pathogens were similar between the 2 treatment groups. Acinetobacter baumannii (66 cases [77%]) was the most common pathogen, followed by Klebsiella pneumoniae (12 cases [14%]) and Pseudomonas aeruginosa (8 cases [9.3%]). No colistin-resistant strains were isolated from patients in either group. No significant differences between the 2 groups were observed regarding eradication of pathogens (P = .679), clinical cure (P = .10), and mortality (P = .289). Eight patients (19%) in each treatment group developed reversible renal dysfunction. No AS colistin-related adverse events were recorded.

Conclusions. Addition of AS colistin to IV colistin did not provide additional therapeutic benefit to patients with MDR VAP due to gram-negative bacteria.

Table 2. Clinical and Bacteriological Outcomes, Mortality, and Adverse Events in Both Treatment Groups

Outcome	No. (%) of patients: IV colistin group (n = 43)	AS-IV colistin group (n = 43)	P
Clinical outcome			
Clinical cure	14 (32.5)	23 (54)	.05
Clinical improvement	12 (28)	9 (21)	.451
Clinical failure	14 (32.5)	7 (16)	.126
Recurrence	3 (7)	4 (9)	>.99
Bacteriological outcome[a]			
Eradication	17 (50)	19 (45)	.679
Persistent	12 (35)	10 (24)	.272
Recurrence	2 (6)	5 (12)	.450
Colonization	3 (9)	8 (19)	.208
Mortality			
All-cause	18 (42)	10 (23)	.066
VAP-related	11 (26)	7 (16)	.289
Adverse events			
Nephrotoxicity	8 (19)	8 (19)	>.99
Neurotoxicity	0	0	

NOTE.AS, aerosolized; IV, intravenous; VAP, ventilator-associated pneumonia

[a] Bacteriological outcome was evaluated in 34 patients in the IV colistin group and in 42 patients in the AS-IV colistin group.

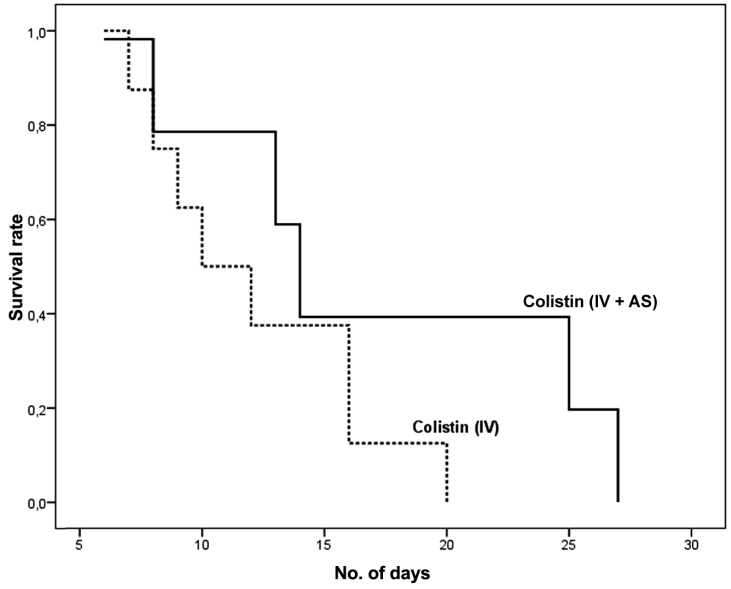


Figure 2. Ventilator-associated pneumonia–related mortality in the 2 treatment groups. AS, aerosolized; IV, intravenous.

this 'Data Management and Statistics' section,

> Data were collected on forms and were computerized and analyzed using SPSS software, version 16.0 (SPSS). Variables for the matched case-control pairs were compared by Wilcoxon matched pairs test. The χ^2 or Fisher exact test was used to assess differences in categorical variables, as appropriate. Differences in continuous variables were assessed by the Student t test or nonparametric Mann-Whitney U test. Kaplan-Meier curves were used to assess differences between the IV group and the IV plus AS group and overall mortality. The log-rank test was used to determine the level of statistical significance when comparing survival curves. Multivariate logistic regression analysis was used to assess the independent effect of therapy on each of the 2 outcomes (clinical cure and microorganism eradication). P values are 2-tailed, and P values $< .05$ were considered to be statistically significant.

as well as this paragraph in the Results

> Overall, the mortality rate in the ICU was 42% (18 of 43 patients) in the IV colistin, compared with and 24% (10 of 43 patients) in the AS-IV colistin group ($P = .066$). The VAP-related mortality rates were 26% (11 of 43 patients) and 16% (7 of 43 patients), respectively ($P = .289$). Kaplan-Meier curves revealed no statistically significant differences in either all-cause mortality ($P = .888$, by log-rank test) or VAP-related mortality ($P = .268$, by log-rank test) (Figures 1 and 2).

prompted a letter to the Editor (link) , which included (among others) the following remarks:

> Curiously, the Kaplan-Meier curves show survival for the deceased persons as they all end at 0% and the numbers of deaths in each group that can be derived from them do not correspond to the numbers given in Table 2.

In their reply to this point, the authors stated that

> Regarding the Kaplan-Meier curves and number of deaths, the numbers in Table 2 are in full accordance with the number of deaths in the curves. The horizontal steps in the curves are step functions, where each step down indicates presence of an event (death in this study). Thus, each death represents a downward step in the curve. When we try to extract the number of events from the curves, it is crucial to keep in mind that two or more events can coexist at a specific time, so the drop can be twice as large or more.

Question for `bios601` students:

1. Comment on the authors' description of their study as 'A Matched Case-Control Study'.
2. Show how the authors arrived at their $P = .066$ for the all-cause mortality comparison. If you used an online calculator, include a screenshot. If you used `R`, show the code you used.
3. Determine how the curves were fitted, and determine the distribution of the times of the 10 and of the 18 deaths.
4. Carry out an analysis that gives the author's $P = .888$ for all-cause mortality.
5. Carry out your version of the log-rank test. Explain any differences between yours and theirs, as well as any assumptions you made.
6. Report on your quick survey of the web/textbooks as to which of the two versions[13] is more common.

[13] The statistic is summed over the 2 x 2 tables for the different risk sets. Let i refer to the ith table. **One version** uses the null hypothesis to calculate 2 separate expected numbers of events, E_{i1} and E_{i2}, and sums these (and the corresponding observed numbers O_{i1} and O_{i2}) over all tables to give an overall O_1 and E_1 and an overall O_2 and E_2.The statistic is then computed as $X^2 = \frac{(O_1-E_1)^2}{E_1} + \frac{(O_2-E_2)^2}{E_2}$.
The other version focuses only on the observed and expected frequencies in one cell (usually the 'a' cell, although it doesn't matter which one cell you choose to focus on). One then sums (over the tables) the excesses or deficits {the values of $a_i - E[a_i|H_0]$}, and squares this overall deficit or excess. The statistic is $X^2 = \frac{[\ \sum_i (a_i - E[a_i|H_0])\ ]^2}{\sum_i V_i}$, where V_i is the variance, calculated under the null, of the random quantity $a_i - E[a_i|H_0]$. Under the hypergeometric distribution, with row, column, and overall totals r_1, r_2, c_1, c_2 and n, it has the form $V_i = \frac{r_1 \times r_2 \times c_1 \times c_2}{n^2(n-1)}$. This version of the statistic is the same one that Mantel and Haenszel proposed in 1959 for stratified tables in case-control studies.
See also: Chapter 17 (Survival Analysis), from Armitage et al. 4th edition, in Resources.

Supplementary Exercise 4.15 (Full electronic article here)

Association Between Push-up Exercise Capacity and Future Cardiovascular Events Among Active Adult Men

Justin Yang, MD, MPH; Costas A. Christophi, PhD; Andrea Farioli, MD, PhD; Dorothee M. Baur, MD, MS; Steven Moffatt, MD; Terrell W. Zollinger, DrPH; Stefanos N. Kales, MD, MPH

Abstract

IMPORTANCE Cardiovascular disease (CVD) remains the leading cause of mortality worldwide. Robust evidence indicates an association of increased physical fitness with a lower risk of CVD events and improved longevity; however, few have studied simple, low-cost measures of functional status.

OBJECTIVE To evaluate the association between push-up capacity and subsequent CVD event incidence in a cohort of active adult men.

DESIGN, SETTING, AND PARTICIPANTS Retrospective longitudinal cohort study conducted between January 1, 2000, and December 31, 2010, in 1 outpatient clinics in Indiana of male firefighters aged 18 years or older. Baseline and periodic physical examinations, including tests of push-up capacity and exercise tolerance, were performed between February 2, 2000, and November 12, 2007. Participants were stratified into 5 groups based on number of push-ups completed and were followed up for 10 years. Final statistical analyses were completed on August 11, 2018.

MAIN OUTCOMES AND MEASURES Cardiovascular disease–related outcomes through 2010 included incident diagnoses of coronary artery disease and other major CVD events. Incidence rate ratios (IRRs) were computed, and logistic regression models were used to model the time to each outcome from baseline, adjusting for age and body mass index (BMI) (calculated as weight in kilograms divided by height in meters squared). Kaplan-Meier estimates for cumulative risk were computed for the push-up categories.

RESULTS A total of 1562 participants underwent baseline examination, and 1104 with available push-up data were included in the final analyses. Mean (SD) age of the cohort at baseline was 39.6 (9.2) years, and mean (SD) BMI was 28.7 (4.3). During the 10-year follow up, 37 CVD-related outcomes (8601 person-years) were reported in participants with available push-up data. Significant negative associations were found between increasing push-up capacity and CVD events. Participants able to complete more than 40 push-ups were associated with a significantly lower risk of incident CVD event risk compared with those completing fewer than 10 push-ups (IRR, 0.04; 95% CI, 0.01-0.36).

CONCLUSIONS AND RELEVANCE The findings suggest that higher baseline push-up capacity is associated with a lower incidence of CVD events. Although larger studies in more diverse cohorts are needed, push-up capacity may be a simple, no-cost measure to estimate functional status.

JAMA Network Open. 2019;2(2):e188341. doi:10.1001/jamanetworkopen.2018.8341

Key Points

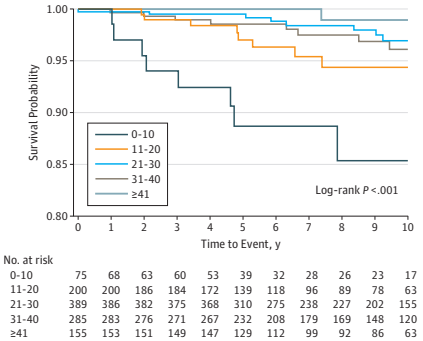
Question Is there an office-based objective measurement that clinicians can use to assess the association between fitness and cardiovascular disease risk?

Findings This longitudinal cohort study of 1104 occupationally active adult men found a significant negative association between baseline push-up capacity and incident cardiovascular disease risk across 10 years of follow-up. Participants able to complete more than 40 push-ups were associated with a significant reduction in incident cardiovascular disease event risk compared with those completing fewer than 10 push-ups.

Meaning Push-up capacity is a no-cost, fast, and simple measure that may be a useful and objective clinical assessment tool for evaluating functional capacity and cardiovascular disease risk.

Author affiliations and article information are listed at the end of this article.

Figure. Kaplan-Meier Curves for the Cumulative Risk of Cardiovascular Disease Outcome in 5 Push-up Categories



Push-up categories are by numbers of push-ups performed during baseline examination.

Table 1. Descriptive Characteristics of Study Participants With Available Push-up Data Stratified by Number of Push-ups Performed During Baseline Examination

	All Participants		0-10 Push-ups		11-20 Push-ups		21-30 Push-ups		31-40 Push-ups		≥41 Push-ups		
Variable	No.	Mean (SD)	No.	Mean (SD)	No.	Mean (SD)	No.	Mean (SD)	No.	Mean (SD)	No.	Mean (SD)	*P* Value[a]
Age, y	1104	39.6 (9.2)	75	48.4 (10.1)	200	45.1 (8.6)	389	39.0 (8.3)	285	36.6 (8.0)	155	35.1 (7.1)	<.001
BMI	1101	28.7 (4.3)	75	33.1 (5.8)	200	30.3 (4.9)	388	28.7 (3.9)	285	27.4 (3.1)	155	26.8 (2.9)	<.001
Blood pressure, mm Hg													
SBP	1104	127.5 (12.0)	75	136.9 (17.9)	200	129.6 (12.1)	389	126.9 (11.8)	285	125.6 (10.3)	155	125.2 (9.4)	<.001
DBP	1104	85.7 (7.9)	75	89.4 (9.5)	200	86.5 (8.4)	389	85.9 (7.7)	285	84.6 (7.5)	155	84.0 (7.2)	<.001
Cholesterol level, mg/dL													
Total	1066	198.3 (38.1)	75	201.7 (43.0)	197	201.5 (35.6)	376	201.3 (39.8)	270	194.8 (37.2)	148	191.0 (34.9)	.02
HDL	1067	47.3 (23.1)	75	41.9 (10.6)	198	45.6 (15.1)	376	47.7 (34.6)	270	48.3 (12.4)	148	49.6 (11.2)	.13
LDL	1030	125.3 (42.0)	71	130.6 (33.3)	190	130.6 (70.8)	363	126.7 (32.3)	262	120.4 (31.4)	144	120.8 (31.3)	.04
Triglycerides	1066	145.2 (109.3)	75	167.9 (99.6)	197	162.2 (113.9)	376	150.9 (112.2)	270	134.5 (109.2)	148	116.1 (92.6)	<.001
Glucose level, mg/dL	1066	88.9 (16.4)	75	99.4 (24.3)	197	93.7 (22.9)	376	88.0 (13.9)	270	86.0 (12.4)	148	85.0 (8.4)	<.001
$\dot{V}O_2$max	1104	43.2 (6.3)	75	37.9 (6.5)	200	41.4 (6.0)	389	43.2 (6.2)	285	44.4 (5.7)	155	45.9 (5.4)	<.001
Race/ethnicity, No. (%)													
White	NA	964 (87.7)	NA	66 (88.0)	NA	170 (85.9)	NA	347 (89.2)	NA	245 (86.2)	NA	136 (88.3)	.95
African American	NA	118 (10.7)	NA	8 (10.7)	NA	25 (12.6)	NA	35 (9.0)	NA	34 (12.0)	NA	16 (10.4)	
Other	NA	18 (1.6)	NA	1 (1.3)	NA	3 (1.5)	NA	7 (1.8)	NA	5 (1.8)	NA	2 (1.3)	
Smoking status, No. (%)													
Nonsmoker	NA	617 (57.7)	NA	34 (45.3)	NA	82 (41.2)	NA	216 (57.4)	NA	182 (67.4)	NA	103 (69.1)	<.001
Previous smoker	NA	295 (27.6)	NA	23 (30.7)	NA	70 (35.2)	NA	102 (27.1)	NA	64 (23.7)	NA	36 (24.2)	
Current smoker	NA	157 (14.7)	NA	18 (24.0)	NA	47 (23.6)	NA	58 (15.4)	NA	24 (8.9)	NA	10 (6.7)	

Abbreviations: BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); DBP, diastolic blood pressure, HDL, high-density lipoprotein; LDL, low-density lipoprotein; NA, not applicable; SBP, systolic blood pressure; $\dot{V}O_2$max, maximal oxygen consumption.

SI conversion: To convert total, HDL, and LDL cholesterol to mmol/L, multiply by 0.0259; triglycerides to mmol/L, multiply by 0.0113; and glucose to mmol/L, multiply by 0.0555.

[a] *P* value based on an analysis of variance or χ^2 test.

1. We will focus on the K-M curves in Figure 1 (and the Incidence Rate Ratios in the top half of Table 2) more to get experience with these entities than to make fair comparisons. *Why, based on the data reported in Table 1, do these 'crude' comparisons over-sell the benefits (for cardiovascular health) of being able to complete more push-ups? (Put another way, why were the authors asked to show Table 3 in addition to Table 2?)*

2. Read the last sentence of the Results section of the Abstract, and identify the table in which this result appears. Comment.

3. From data in the Figure, construct a dataset of 1104 observations (1 per participant) that comes close to the actual dataset, and use it – and the `survival` package in `R` – to generate the K-M curves. It will help to work

with the electronic version of the Figure, so that you can enlarge it.

4. For a (crude) log-rank test involving just the 11-20 (index category) versus the 0-10 (reference category) comparison, show the calculations involved in the contributions from the 3rd, 4th, and last risksets. Different ways of calculating the log-rank statistic are given in footnote 14 for exercise 4.14; on pages 146, 147 and 151 of section 15.4 of Chapter 15 of Clayton and Hills (link); and on page 4 of these Notes link). The example in Clayton and Hills uses a finer time-scale, and so there is just 1 event per riskset. The example in the Notes uses a coarser time-scale, and so some of the random variables have null Binomial (or hypergeometric) distributions, rather than Bernoulli ones.

odds model with an eight-category ordinal scale, the patients who received remdesivir were found to be more likely than those who received placebo to have clinical improvement at day 15 (odds ratio, 1.5; 95% CI, 1.2 to 1.9, after adjustment for actual disease severity). The Kaplan–Meier estimates of mortality were 6.7% with remdesivir and 11.9% with placebo by day 15 and 11.4% with remdesivir and 15.2% with placebo by day 29 (hazard ratio, 0.73; 95% CI, 0.52 to 1.03). Serious adverse events were reported in 131 of the 532 patients who received remdesivir (24.6%) and in 163 of the 516 patients who received placebo (31.6%).

CONCLUSIONS

Our data show that remdesivir was superior to placebo in shortening the time to recovery in adults who were hospitalized with Covid-19 and had evidence of lower respiratory tract infection. (Funded by the National Institute of Allergy and Infectious Diseases and others; ACTT-1 ClinicalTrials.gov number, NCT04280705.)

Table 3. Comparison Between Multiple Models of the Association of Maximal Oxygen Consumption or Push-up Categories With Cardiovascular Disease Outcome[a]

Model	HR (95% CI) Adjusted for Age[b]	*P* Value	HR (95% CI) Adjusted for Age and BMI[c]	*P* Value
Model 1 ($\dot{V}o_2$max)				
5 vs 1	0.52 (0.05-5.16)	.58	0.56 (0.05-5.90)	.63
4 vs 1	1.51 (0.40-5.76)	.54	1.60 (0.38-6.67)	.52
3 vs 1	0.75 (0.19-3.00)	.69	0.81 (0.18-3.71)	.78
2 vs 1	0.53 (0.15-1.85)	.32	0.56 (0.15-2.06)	.38
Model 2 (Push-up Categories)[b]				
5 vs 1	0.15 (0.02-1.29)	.08	0.14 (0.02-1.22)	.07
4 vs 1	0.60 (0.21-1.67)	.32	0.53 (0.17-1.66)	.28
3 vs 1	0.33 (0.12-0.90)	.03	0.31 (0.11-0.89)	.03
2 vs 1	0.47 (0.18-1.23)	.12	0.45 (0.17-1.20)	.11
Model 3 ($\dot{V}o_2$max and Push-up Categories)[b]				
$\dot{V}o_2$max				
5 vs 1	0.63 (0.06-6.38)	.69	0.54 (0.05-5.89)	.61
4 vs 1	2.09 (0.52-8.48)	.30	1.82 (0.41-8.10)	.43
3 vs 1	0.89 (0.22-3.66)	.87	0.74 (0.15-3.60)	.71
2 vs 1	0.64 (0.18-2.32)	.50	0.57 (0.15-2.23)	.42
Push-up categories				
5 vs 1	0.13 (0.01-1.14)	.07	0.11(0.01-1.07)	.06
4 vs 1	0.52 (0.17-1.54)	.23	0.46 (0.14-1.49)	.20
3 vs 1	0.27 (0.09-0.78)	.02	0.25 (0.08-0.76)	.01
2 vs 1	0.43(0.16-1.19)	.10	0.42 (0.15-1.15)	.09

Abbreviations: BMI, body mass index (calculated as weight in kilograms divided by height in meters squared); HR: hazard ratio; $\dot{V}o_2$max, maximal oxygen consumption.

[a] Push-up categories are defined as follows: category 1, 0 to 10 push-ups; category 2, 11 to 20 push-ups; category 3, 21 to 30 push-ups; category 4, 31 to 40 push-ups; and category 5, 41 push-ups or more. Cardiovascular disease outcome was defined as cardiovascular events including diagnoses of coronary artery disease, or other major cardiovascular disease event and included 37 events per 8601 person-years among 1104 participants.

[b] Adjusted for age using the Cox proportional hazards model.

[c] Adjusted for age and BMI using the Cox proportional hazards model.

5. Just using the data in the Figure, determine the (approx.) numbers of person years in each of the 5 push-up categories, and compare them with those back-calculated from Table 2.

6. Calculate a crude IRR and 95% CI for the IRR for the 11-20 (index category) versus the 0-10 (reference category) contrast. Compare them with those reported in Table 2. [*Hint: work with* $\log \widehat{IRR}$*, so that its variance is the sum of the variances of the logs of the 2 Poisson random variables – encountered already in exercise 0.1 in the notes on intensity rates:- models / inference / planning ; compute the CI in the log scale, then transfer it back to the IRR scale.*]

7. Why are the corresponding *adjusted* IRR (2 vs 1) estimates from model 2 in Table 3 closer to the null (i.e., to IRR=1) than the crude one in Table 2?

Remdesivir for the Treatment of Covid-19 — Preliminary Report

J.H. Beigel, K.M. Tomashek, L.E. Dodd, A.K. Mehta, B.S. Zingman, A.C. Kalil, E. Hohmann, H.Y. Chu, A. Luetkemeyer, S. Kline, D. Lopez de Castilla, R.W. Finberg, K. Dierberg, V. Tapson, L. Hsieh, T.F. Patterson, R. Paredes, D.A. Sweeney, W.R. Short, G. Touloumi, D.C. Lye, N. Ohmagari, M. Oh, G.M. Ruiz-Palacios, T. Benfield, G. Fätkenheuer, M.G. Kortepeter, R.L. Atmar, C.B. Creech, J. Lundgren, A.G. Babiker, S. Pett, J.D. Neaton, T.H. Burgess, T. Bonnett, M. Green, M. Makowski, A. Osinusi, S. Nayak, and H.C. Lane, for the ACTT-1 Study Group Members*

ABSTRACT

BACKGROUND

Although several therapeutic agents have been evaluated for the treatment of coronavirus disease 2019 (Covid-19), none have yet been shown to be efficacious.

METHODS

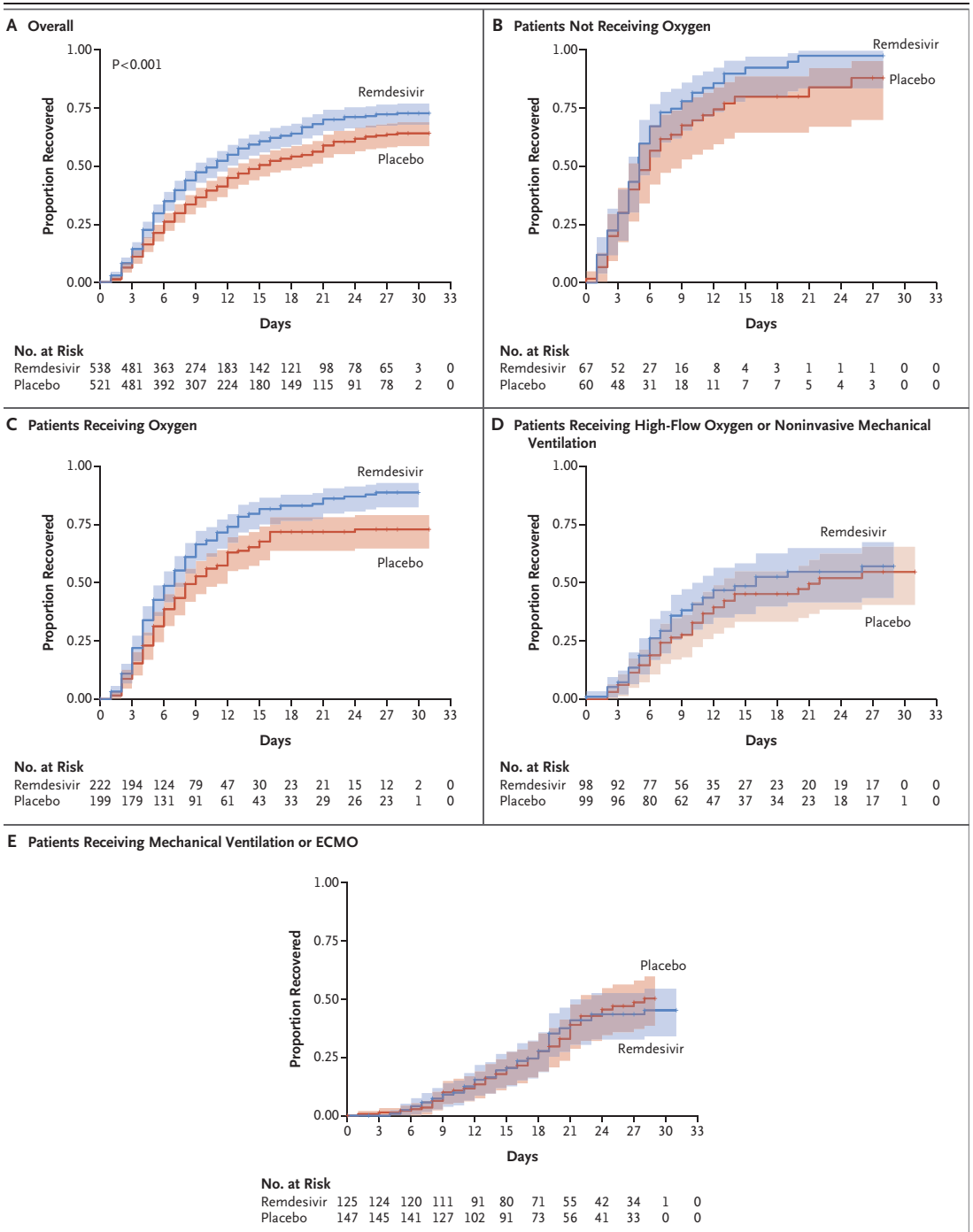
We conducted a double-blind, randomized, placebo-controlled trial of intravenous remdesivir in adults hospitalized with Covid-19 with evidence of lower respiratory tract involvement. Patients were randomly assigned to receive either remdesivir (200 mg loading dose on day 1, followed by 100 mg daily for up to 9 additional days) or placebo for up to 10 days. The primary outcome was the time to recovery, defined by either discharge from the hospital or hospitalization for infection-control purposes only.

RESULTS

A total of 1063 patients underwent randomization. The data and safety monitoring board recommended early unblinding of the results on the basis of findings from an analysis that showed shortened time to recovery in the remdesivir group. Preliminary results from the 1059 patients (538 assigned to remdesivir and 521 to placebo) with data available after randomization indicated that those who received remdesivir had a median recovery time of 11 days (95% confidence interval [CI], 9 to 12), as compared with 15 days (95% CI, 13 to 19) in those who received placebo (rate ratio for recovery, 1.32; 95% CI, 1.12 to 1.55; P<0.001). The Kaplan-Meier estimates of mortality by 14 days were 7.1% with remdesivir and 11.9% with placebo (hazard ratio for death, 0.70; 95% CI, 0.47 to 1.04). Serious adverse events were reported for 114 of the 541 patients in the remdesivir group who underwent randomization (21.1%) and 141 of the 522 patients in the placebo group who underwent randomization (27.0%).

CONCLUSIONS

Remdesivir was superior to placebo in shortening the time to recovery in adults hospitalized with Covid-19 and evidence of lower respiratory tract infection. (Funded by the National Institute of Allergy and Infectious Diseases and others; ACCT-1 ClinicalTrials.gov number, NCT04280705.)



Remdesivir for the Treatment of Covid-19 — Final Report

J.H. Beigel, K.M. Tomashek, L.E. Dodd, A.K. Mehta, B.S. Zingman, A.C. Kalil, E. Hohmann, H.Y. Chu, A. Luetkemeyer, S. Kline, D. Lopez de Castilla, R.W. Finberg, K. Dierberg, V. Tapson, L. Hsieh, T.F. Patterson, R. Paredes, D.A. Sweeney, W.R. Short, G. Touloumi, D.C. Lye, N. Ohmagari, M. Oh, G.M. Ruiz-Palacios, T. Benfield, G. Fätkenheuer, M.G. Kortepeter, R.L. Atmar, C.B. Creech, J. Lundgren, A.G. Babiker, S. Pett, J.D. Neaton, T.H. Burgess, T. Bonnett, M. Green, M. Makowski, A. Osinusi, S. Nayak, and H.C. Lane, for the ACTT-1 Study Group Members*

ABSTRACT

BACKGROUND

Although several therapeutic agents have been evaluated for the treatment of coronavirus disease 2019 (Covid-19), no antiviral agents have yet been shown to be efficacious.

METHODS

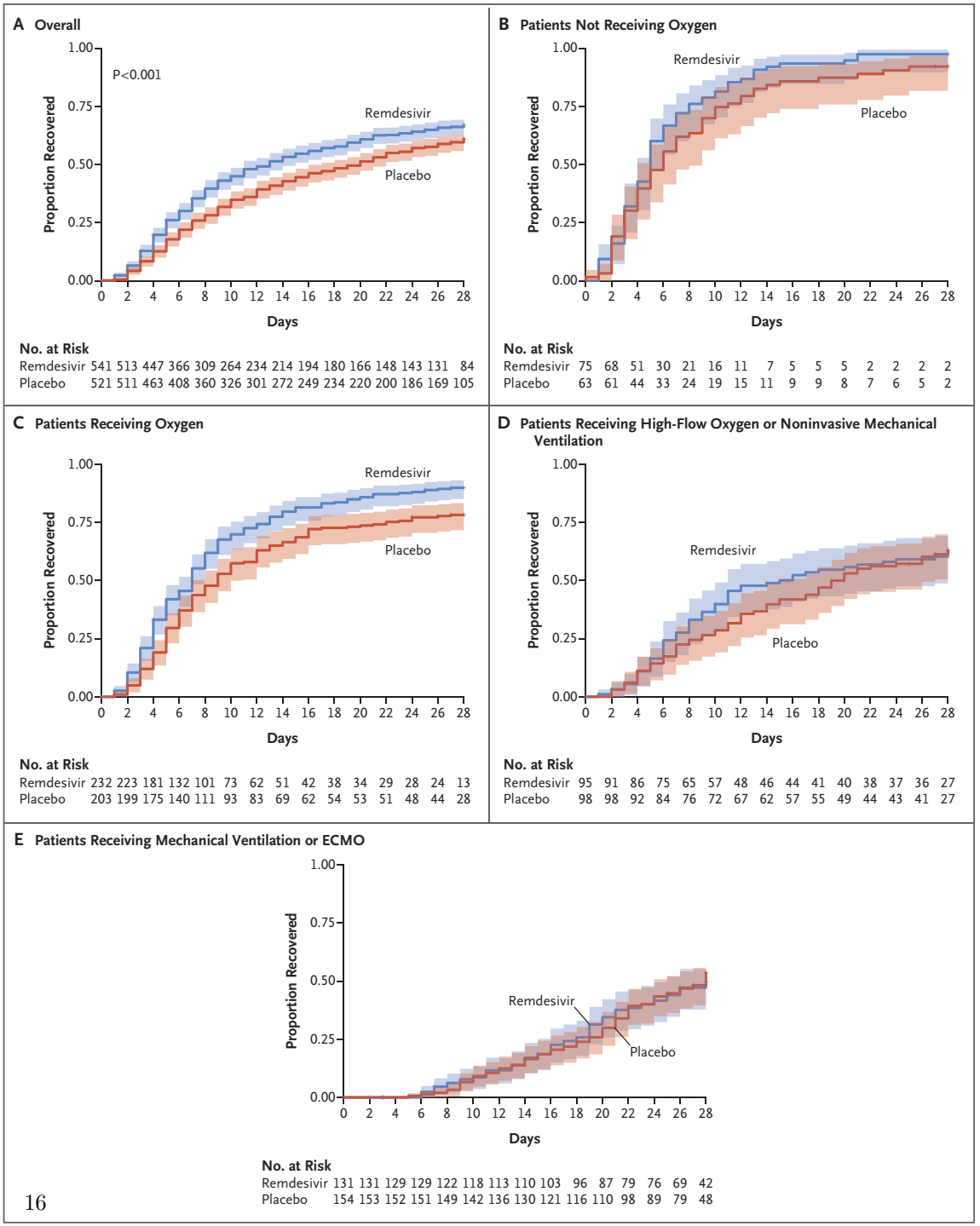
We conducted a double-blind, randomized, placebo-controlled trial of intravenous remdesivir in adults who were hospitalized with Covid-19 and had evidence of lower respiratory tract infection. Patients were randomly assigned to receive either remdesivir (200 mg loading dose on day 1, followed by 100 mg daily for up to 9 additional days) or placebo for up to 10 days. The primary outcome was the time to recovery, defined by either discharge from the hospital or hospitalization for infection-control purposes only.

RESULTS

A total of 1062 patients underwent randomization (with 541 assigned to remdesivir and 521 to placebo). Those who received remdesivir had a median recovery time of 10 days (95% confidence interval [CI], 9 to 11), as compared with 15 days (95% CI, 13 to 18) among those who received placebo (rate ratio for recovery, 1.29; 95% CI, 1.12 to 1.49; P<0.001, by a log-rank test). In an analysis that used a proportional-odds model with an eight-category ordinal scale, the patients who received remdesivir were found to be more likely than those who received placebo to have clinical improvement at day 15 (odds ratio, 1.5; 95% CI, 1.2 to 1.9, after adjustment for actual disease severity). The Kaplan–Meier estimates of mortality were 6.7% with remdesivir and 11.9% with placebo by day 15 and 11.4% with remdesivir and 15.2% with placebo by day 29 (hazard ratio, 0.73; 95% CI, 0.52 to 1.03). Serious adverse events were reported in 131 of the 532 patients who received remdesivir (24.6%) and in 163 of the 516 patients who received placebo (31.6%).

CONCLUSIONS

Our data show that remdesivir was superior to placebo in shortening the time to recovery in adults who were hospitalized with Covid-19 and had evidence of lower respiratory tract infection. (Funded by the National Institute of Allergy and Infectious Diseases and others; ACTT-1 ClinicalTrials.gov number, NCT04280705.)



Supplementary Exercise 4.16

The full articles on the ACCT1 (Rendesivir) trial, as well as a Supplement that includes an expanded Statistical Analysis Plan, can be found in this single file.

Tue **2020-06-30** 9:04 PM.

Hi Jim. Hope all is well and you are surviving these crazy times.

I seem to recall you said at some point in the past that you were able to digitize pdf graphs

Basically for Figure A on page 6 I want to calculate the AUC between the 2 curves as this will give the extra number of recovery days gained with the intervention. I think this is a more useful measure than giving the OR for recovery at arbitrary time points.

Do you think that concept is reasonable. If so, are you able to calculate this area between the curves?

Cheers ..... Jay

Wed **2020-07-01** 2:54 PM Thanks Jim!

This confirms my gut instinct that while the relative metrics in the paper suggest a large benefit, when you look at absolute metrics, the benefit appear smaller.

Quickly looking at your digital plot, your calculations seem right. Each square represents 1 day and 5% difference. I quickly counted about 50 squares between the 2 curves so 50*0.05 = 2.5 people days which approximates your calculations.

This is less than the reported median difference of 4 days which I feel is an exaggeration of the true effect size. Not quite sure how to explain other than comparing than the benefits of examining the whole distributions versus looking at 1 time point.

In fact, I don't believe the choice of median times was mentioned as either a primary or secondary outcome. "The primary outcome measure was the time to recovery, defined as the first day, during the 28 days after enrollment," Moreover this trial suffers from enormous lost to follow-up if 28 days was the endpoint, ignoring deaths, it looks like 90% didn't reach the specified follow-up of 28 days. Maybe those missing people would have further shrunk the differences.

Like your `R` program. I see you haven't been swept up with the tidyverse / ggplot2 universe.

Interestingly about 2 hours ago, BMJ asked me to write an opinion piece about this Guardian article this Guardian article. Eventually we should do a formal cost-effectiveness piece on this drug - although it could be argued that it is a no-brainer in a public system to stay away from it and let the Americans over spend for these very modest benefits. I'll get back to you on this. Cheers

Tue **2020-10-13** 8:10 PM Hi Jim

These exercises look great. Wish I was back being a stats student!

Nice to see the reference to Clayton and Hill, I still have their textbook which remains among my favourites. Reminds me of a statistical epidemiology summer course i took from David Clayton many years ago in Florence. We had some intense ping pong games in the evenings!

So for remdesivir the opinion piece i wrote for the BMJ is found here

Big study apparently to be published this week will confirm no mortality benefit with remdesivir so another reason besides the cost not to be rushing out to be first in line to spend our limited health dollars on this particular drug. Glad for you to reference my email or anything else you think useful.

Stay healthy. Jay website: www.brophyj.com. twitter: @brophyj

James (Jay) Brophy MD PhD Professor of Medicine & Epidemiology (McGill University)

https://www.rte.ie/brainstorm/2020/1013/1171221-remdesivir-magic-bullet-covid-19-donald-trump-tests/

Questions

1. Using the information in the Figure of the 'Remdesivir for the Treatment of Covid-19 — Preliminary Report' carry out the computation Dr Brophy proposed. [See JH for details on extracting data from K-M type curves in pdf files, as well as the article Recovering the raw data behind a non-parametric survival curve] and some `R` code to extract graph co-ordinates from a PostScript file.]
2. Suggest a way to calculate a CI for the area between the curves.
3. In the 'Additional Statistical Analysis Details' section of the 'Supplementary Appendix to Manuscript Entitled Remdesivir for the Treatment of COVID-19 – Final Report' we read

   > The primary analysis was a log-rank test of time-to-recovery between remdesivir and placebo stratified by disease severity as defined above.

   Carry out the log-rank test.
4. We also read

   > The relevant treatment efficacy parameter is the "recovery rate ratio" (for remdesivir relative to placebo), which is akin to the hazard ratio in survival analysis but for the beneficial outcome of recovery.[14] The study was designed to achieve 85% power for detecting a recovery rate ratio of 1.35 with a two-sided type-I error rate of 5%. Enrollment continued through April 19, 2020 to ensure at least 400 recoveries and to address subgroup analysis.

   Carry out the sample size calculations (focusing on a minimum number of recoveries) based on (a) a binomial test that fixes the total number of recovered patients (as in the Mayo Lung Screening trial) and (b) the log of the recovery rate ratio; its variance is $1/E[n.r_0] + 1/E[n.r_1]$, where $n.r_0$ and $n.r_1$ are the numbers of recovered patients in the placebo and remdesivir arms respectively.

[14] "Two practical considerations result from considering time to a beneficial outcome. First, a recovery rate ratio greater than one indicates an improvement for remdesivir. Second, failure to recover and death are both censored at Day 29. Consequently, participants censored on the last observation day reflect two different states: death and failure to recover by Day 29. Hence, a breakdown of deaths by treatment arm is also important to understanding treatment efficacy. The key secondary analysis tested a difference in the ordinal score distribution between remdesivir and placebo at Day 15 using the "common odds ratio" from a proportional odds model, stratifying by baseline disease severity stratum."

Dexamethasone in Hospitalized Patients with Covid-19 — Preliminary Report

The RECOVERY Collaborative Group*

ABSTRACT

BACKGROUND

Coronavirus disease 2019 (Covid-19) is associated with diffuse lung damage. Glucocorticoids may modulate inflammation-mediated lung injury and thereby reduce progression to respiratory failure and death.

METHODS

In this controlled, open-label trial comparing a range of possible treatments in patients who were hospitalized with Covid-19, we randomly assigned patients to receive oral or intravenous dexamethasone (at a dose of 6 mg once daily) for up to 10 days or to receive usual care alone. The primary outcome was 28-day mortality. Here, we report the preliminary results of this comparison.

RESULTS

A total of 2104 patients were assigned to receive dexamethasone and 4321 to receive usual care. Overall, 482 patients (22.9%) in the dexamethasone group and 1110 patients (25.7%) in the usual care group died within 28 days after randomization (age-adjusted rate ratio, 0.83; 95% confidence interval [CI], 0.75 to 0.93; P<0.001). The proportional and absolute between-group differences in mortality varied considerably according to the level of respiratory support that the patients were receiving at the time of randomization. In the dexamethasone group, the incidence of death was lower than that in the usual care group among patients receiving invasive mechanical ventilation (29.3% vs. 41.4%; rate ratio, 0.64; 95% CI, 0.51 to 0.81) and among those receiving oxygen without invasive mechanical ventilation (23.3% vs. 26.2%; rate ratio, 0.82; 95% CI, 0.72 to 0.94) but not among those who were receiving no respiratory support at randomization (17.8% vs. 14.0%; rate ratio, 1.19; 95% CI, 0.91 to 1.55).

CONCLUSIONS

In patients hospitalized with Covid-19, the use of dexamethasone resulted in lower 28-day mortality among those who were receiving either invasive mechanical ventilation or oxygen alone at randomization but not among those receiving no respiratory support. (Funded by the Medical Research Council and National Institute for Health Research and others; RECOVERY ClinicalTrials.gov number, NCT04381936; ISRCTN number, 50189673.)

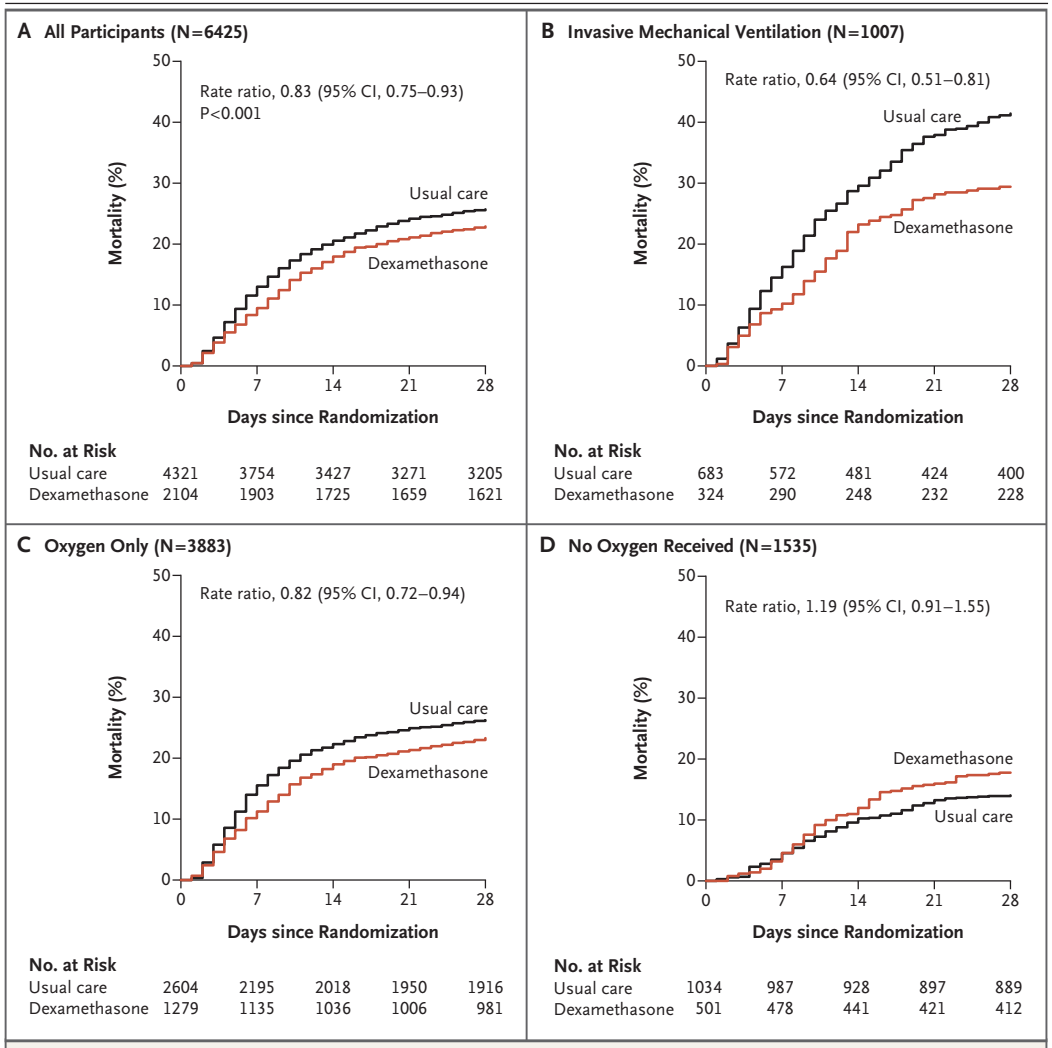


Figure 2. Mortality at 28 Days in All Patients and According to Respiratory Support at Randomization.

Shown are Kaplan–Meier survival curves for 28-day mortality among all the patients in the trial (primary outcome) (Panel A) and in three respiratory-support subgroups according to whether the patients were undergoing invasive mechanical ventilation (Panel B), receiving oxygen only without mechanical ventilation (Panel C), or receiving no supplemental oxygen (Panel D) at the time of randomization. The Kaplan–Meier curves have not been adjusted for age. The rate ratios have been adjusted for the age of the patients in three categories (<70 years, 70 to 79 years, and ≥80 years). Estimates of the rate ratios and 95% confidence intervals in Panels B, C, and D were derived from a single age-adjusted regression model involving an interaction term between treatment assignment and level of respiratory support at randomization.

Supplementary Exercise 4.17

The full article Dexamethasone in Hospitalized Patients with Covid-19 – Preliminary Report is found here.

1. The Statistical Analysis section begins...

   > As stated in the protocol, appropriate sample sizes could not be estimated when the trial was being planned at the start of the Covid-19 pandemic. As the trial progressed, the trial steering committee, whose members were unaware of the results of the trial comparisons, determined that if 28-day mortality was 20%, then the enrollment of at least 2000 patients in the dexamethasone group and 4000 in the usual care group would provide a power of at least 90% at a two-sided P value of 0.01 to detect a clinically relevant proportional reduction of 20% (an absolute difference of 4 percentage points) between the two groups. Consequently, on June 8, 2020, the steering committee closed recruitment to the dexa-methasone group, since enrollment had exceed- ed 2000 patients.

   Do your own power/sample size calculations and compare them with those above. State any assumptions you made.

2. Repeat the calculations for a design in which, rather than 1:2, the randomization was (a) 1:1 (b) 1:3. Comment on the lessons you learned from these calculations.

3. The section went on to say

   > For the primary outcome of 28-day mortality, the hazard ratio from Cox regression was used to estimate the mortality rate ratio. Among the few patients (0.1%) who had not been followed for 28 days by the time of the data cutoff on July 6, 2020, data were censored either on that date or on day 29 if the patient had already been discharged. That is, in the absence of any information to the contrary, these patients were assumed to have survived for 28 days. Kaplan–Meier survival curves were constructed to show cumulative mortality over the 28-day period.

4. Use the numbers in the Figure to verify that the censoring was indeed minimal and negligible.

5. How does this information simplify the calculation of the SE for the difference in 28-day mortality rates?

6. Calculate a 95% CI for ratio of the 28-day mortality rates (unlike the authors, you don't have the data to calculate the age-adjusted ratio.)

7. Is the ratio in patients receiving invasive mechanical ventilation significantly different from the ratio in those receiving oxygen without invasive mechanical ventilation?

8. For each of these two classes of patients, calculate the number needed to treat to prevent one death, and try to find the 'costs' of doing so. See the Dr Brophy's BMJ blog for the cost calculations for Remdesivir.

9. Use this trial to explain why, for doctors, knowing when there is effect modification (different slopes – or different effects – for different folks, or 'interaction' to statisticians) is very important. 'Interactions' make in statistical models more complex, and the story more nuanced; one answer doesn't fit all, rather 'it depends'. But 'le bon traitement pour le bon patient' is central to good medical practice.

Effect of Hydroxychloroquine in Hospitalized Patients with Covid-19

The RECOVERY Collaborative Group*

ABSTRACT

BACKGROUND

Hydroxychloroquine and chloroquine have been proposed as treatments for coronavirus disease 2019 (Covid-19) on the basis of in vitro activity and data from uncontrolled studies and small, randomized trials.

METHODS

In this randomized, controlled, open-label platform trial comparing a range of possible treatments with usual care in patients hospitalized with Covid-19, we randomly assigned 1561 patients to receive hydroxychloroquine and 3155 to receive usual care. The primary outcome was 28-day mortality.

RESULTS

The enrollment of patients in the hydroxychloroquine group was closed on June 5, 2020, after an interim analysis determined that there was a lack of efficacy. Death within 28 days occurred in 421 patients (27.0%) in the hydroxychloroquine group and in 790 (25.0%) in the usual-care group (rate ratio, 1.09; 95% confidence interval [CI], 0.97 to 1.23; P=0.15). Consistent results were seen in all prespecified subgroups of patients. The results suggest that patients in the hydroxychloroquine group were less likely to be discharged from the hospital alive within 28 days than those in the usual-care group (59.6% vs. 62.9%; rate ratio, 0.90; 95% CI, 0.83 to 0.98). Among the patients who were not undergoing mechanical ventilation at baseline, those in the hydroxychloroquine group had a higher frequency of invasive mechanical ventilation or death (30.7% vs. 26.9%; risk ratio, 1.14; 95% CI, 1.03 to 1.27). There was a small numerical excess of cardiac deaths (0.4 percentage points) but no difference in the incidence of new major cardiac arrhythmia among the patients who received hydroxychloroquine.

CONCLUSIONS

Among patients hospitalized with Covid-19, those who received hydroxychloroquine did not have a lower incidence of death at 28 days than those who received usual care. (Funded by UK Research and Innovation and National Institute for Health Research and others; RECOVERY ISRCTN number, ISRCTN50189673; ClinicalTrials .gov number, NCT04381936.)

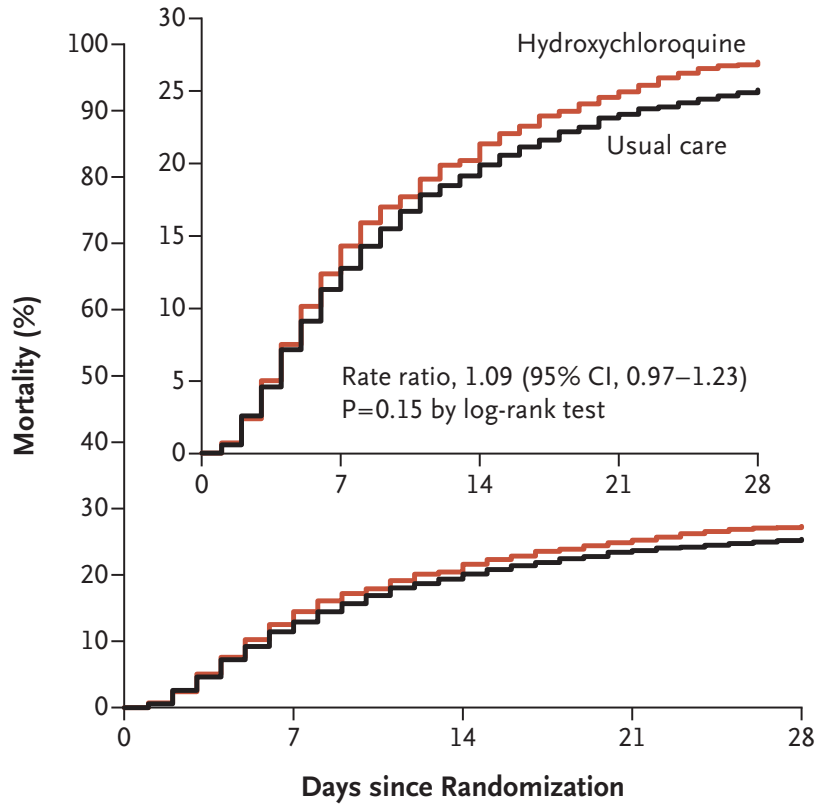


Figure 2. Mortality at 28 Days.

Death at 28 days (the primary outcome) occurred in 421 patients (27.0%) in the hydroxychloroquine group and in 790 (25.0%) in the usual-care group. The inset shows the same data on an expanded y axis.

Supplementary Exercise 4.18

RANDOMISED EVALUATION OF COVID-19 THERAPY(RECOVERY)

This national clinical trial aims to identify treatments that may be beneficial for people hospitalised with suspected or confirmed COVID-19

The full article Effect of Hydroxychloroquine Hospitalized Patients with Covid-19 is found here.

OUTCOME MEASURES

The primary outcome was all-cause mortality within 28 days after randomization; further analyses were specified at 6 months. Secondary outcomes were the time until discharge from the hospital and a composite of the initiation of invasive mechanical ventilation including extracorporeal membrane oxygenation or death among patients who were not receiving invasive mechanical ventilation at the time of randomization

STATISTICAL ANALYSIS

For the primary outcome of 28-day mortality, we used the log-rank observed-minus-expected statistic and its variance both to test the null hypothesis of equal survival curves and to calculate the one-step estimate of the average mortality rate ratio in the comparison between the hydroxy-chloroquine group and the usual-care group. Kaplan-Meier survival curves were constructed to show cumulative mortality over the 28-day period. The same methods were used to analyze the time until hospital discharge, with censor- ing of data on day 29 for patients who had died in the hospital. We used the Kaplan–Meier estimates to calculate the median time until hospital discharge. For the pre-specified composite secondary outcome of invasive mechanical ventilation or death within 28 days (among patients who had not been receiving invasive mechanical ventilation at randomization), the precise date of the initiation of invasive mechanical ventilation was not available, so the risk ratio was estimated instead. Estimates of the between-group difference in absolute risk were also calculated.

All the analyses were performed according to the intention-to-treat principle. Prespecified analyses of the primary outcome were performed in six subgroups, as defined by characteristics at randomization: age, sex, race, level of respiratory support, days since symptom onset, and predicted 28-day risk of death. (Details are provided in the Supplementary Appendix.)

Estimates of rate and risk ratios are shown with 95% confidence intervals without adjust-ment for multiple testing. The P value for the assessment of the primary outcome is two-sided. The full database is held by the trial team, which collected the data from the trial sites and per- formed the analyses, at the Nuffield Department of Population Health at the University of Oxford.

SUPPLEMENTARY STATISTICAL METHODS

Sample size

As stated in the protocol, appropriate sample sizes could not be estimated when the trial was being planned at the start of the COVID-19 pandemic. As the trial progressed, the Trial Steering Committee, blinded to the results of the study treatment comparisons, formed the view that if 28-day mortality was 20% then a comparison of at least 2000 patients allocated to active drug and 4000 to usual care alone would yield at least 90% power at two-sided P=0.01 to detect a proportional reduction of one-fifth (a clinically relevant absolute difference of 4 percentage points between the two arms).

Baseline-predicted risk

Baseline-predicted risk of 28-day mortality was estimated through the formula $100 \times \exp(a)/(1+\exp(a))$, where $a =$

```
-1.23
- 2.85 (age < 50) - 2.03 (age 50 - 59) - 1.21 (age 60 - 69) - 0.51 (age 70 - 79)
+ 0.42 (male)
- 0.34 (> 7 days since symptom onset)
+ 0.86 (on oxygen only at randomization)
+ 2.18 (on invasive mechanical ventilation at randomization)
- 0.01 (history of diabetes)
+ 0.22 (history of heart disease)
+ 0.21 (history of chronic lung disease)
+ 0.50 (history of kidney disease).
```

These regression coefficients were derived from a multivariable logistic regression model using data from all trial participants who (at the time of data-lock) had complete 28-day mortality follow-up data. The regression model additionally adjusted for treatment allocation (with usual care designated the reference category) and for all possible two-way interactions between the above baseline characteristics and treatment allocation. These additional terms were ignored when calculating baseline-predicted risk, however, in order to ensure that the estimates corresponded to risk if assigned usual care. Patients were then subdivided into three approximately equally-sized groups (across all RECOVERY participants) on the basis of their predicted risk: $< 30\%$, $\geq 30\%$ to $< 45\%$, and $\geq 45\%$. Calculation of rate ratio The RR is derived from the log-rank observed minus expected statistic (O - E) and its variance (V) as the one-step estimate, through the formula exp([O - E] ÷ V), and its 95% CI is given by $\exp([O-E] \div V \pm 1.96 \div V^{1/2})$. simulations were performed and presented as median values and 95% prediction intervals.

Ascertainment and classification of study outcomes Information on baseline

characteristics and study outcomes was collected through a combination of electronic case report forms (see below) completed by members of the local research team at each participating hospital and linkage to National Health Service, clinical audit, and other relevant health records. Full details are provided in the RECOVERY Definition and Derivation of Baseline Characteristics and Outcomes Document which was published online (www.recoverytrial.net) on 9 June 2020. Randomization form The Randomization form (shown below) was completed by trained study staff. It collected baseline information about the participant (including demographics, COVID-19 history, comorbidities and suitability for the study treatments) and availability of the study treatments. Once completed and electronically signed, the treatment allocation was displayed.

RECOVERY
Randomised Evaluation of COVID-19 Therapy

Hydroxychloroquine for COVID-19

Test version only (v6.08 - 05/06/20)

Randomisation Program

Call Freefone **0800 138 5451** to contact the RECOVERY team for **URGENT** problems using the Randomisation Program or for medical advice
All **NON-URGENT queries** should be emailed to **recoverytrial@ndph.ox.ac.uk**

Logged in as: **Barts Health NHS Trust**

Section A: Baseline and Eligibility

Date and time of randomisation: 5 Jun 2020 13:32

Treating clinician

A1. Name of treating clinician

Patient details

A2. Patient surname

Patient forename

A3. NHS number ☐ Tick if not available

A4. What is the patient's date of birth? / /

A5. What is the patient's sex?

Inclusion criteria

A6. Has consent been taken in line with the protocol?
If answer is No patient cannot be enrolled in the study

A7. Does the patient have proven or suspected SARS-CoV-2 infection?
If answer is No patient cannot be enrolled in the study

A8. Does the patient have any medical history that might, in the opinion of the attending clinician, put the patient at significant risk if they were to participate in the trial?

A8B. Is the patient willing to receive convalescent plasma?

A9. COVID-19 symptom onset date: / /

A10. Date of hospitalisation: / /

A11. Does the patient require oxygen?

A12. Does the patient **CURRENTLY** require ventilation or ECMO?
Invasive mechanical ventilation or extra-corporeal membrane oxygenation

Does the patient have any CURRENT comorbidities or other medical problems?

A13.1 Diabetes

A13.2 Heart disease

A13.3 Chronic lung disease

A13.4 Tuberculosis

A13.5 HIV

A13.6 Severe liver disease

A13.7 Severe kidney impairment (eGFR<30 or on dialysis)

A13.8 Known long QT syndrome

A13.9 Current treatment with macrolide antibiotics which are to continue
Macrolide antibiotics include clarithromycin, azithromycin and erythromycin

A13.10 Previous adverse reaction to blood or blood product transfusion

Are the following treatments UNSUITABLE for the patient?
If you answer Yes it means you think this participant should NOT receive this drug.

A14.1 Lopinavir-Ritonavir

A14.3 Azithromycin

A14B.1 Convalescent plasma

Are the following treatments available?

A15.1 Lopinavir-Ritonavir

A15.3 Azithromycin

A15B.1 Convalescent plasma

Current medication

A16 Is the patient currently prescribed remdesivir?

Please sign off this form once complete

Surname:

Forename:

Professional email:

Continue

Cancel

Page 25 of 37

Home

Follow-up form

The Follow-up form collected information on study treatment adherence (including both the randomised allocation and use of other study treatments), vital status (including date and provisional cause of death if available), hospitalisation status (including date of discharge), respiratory support received during the hospitalisation, occurrence of any major cardiac arrhythmias and renal replacement therapy received.

28/05/2020 Follow-up

Hydroxychloroquine for COVID-19

Follow-up

Date of randomisation

Patient's date of birth

yyyy-mm-dd

1. Which of following treatment(s) did the patient definitely receive as part of their hospital admission after randomisation? *

(NB Include RECOVERY study-allocated drug, only if given, PLUS any of the other treatments if given as standard hospital care)

- [ ] No additional treatment
- [ ] Lopinavir-ritonavir
- [ ] Corticosteroid (dexamethasone, prednisolone or hydrocortisone)
- [ ] Hydroxychloroquine
- [ ] Azithromycin or other macrolide (eg, clarithromycin, erythromycin)
- [ ] Tocilizumab or sarilumab
- [ ] Remdesivir

The following questions only appear if the treatments have been allocated at randomisation

Please select number of days the patient received lopinavir-ritonavir

1 2 3 4 5 6 7 8 9 10

Please select number of days the patient received corticosteroid (dexamethasone, prednisolone or hydrocortisone)

1 2 3 4 5 6 7 8 9 10

Please select number of days the patient received hydroxychloroquine

1 2 3 4 5 6 7 8 9 10

Please select number of days the patient received azithromycin **This question and the following question cannot both be zero**

0 1 2 3 4 5 6 7 8 9 10

Please select number of days the patient received other macrolides (eg, clarithromycin, erythromycin)

0 1 2 3 4 5 6 7 8 9 10

Please select number of doses of tocilizumab or sarilumab the patient received

1 >1

28/05/2020 Follow-up

Hydroxychloroquine for COVID-19

Please select number of days the patient received remdesivir

1 2 3 4 5 6 7 8 9 10

» Convalescent Plasma

How many convalescent plasma infusions did the patient receive?

This is plasma given as part of trial, not any standard fresh frozen plasma or other blood products that the patient may have been given

0 1 2

Were any infusions stopped early for any reason ie, the patient did not receive the full amount?

Yes No

How many were stopped early?

1 2

» Health Status

2. Was a COVID-19 test done for this patient?

(If multiple tests were done, and the results were positive and negative, please tick Yes – positive result and Yes – negative result)

- [ ] Yes – positive result
- [ ] Yes – negative result
- [ ] Not done

3. What is the patient's vital status? *

- ◯ Alive
- ◯ Dead

3.1 What is the patient's current hospitalisation status?

Q3.1 is only completed if the patients is alive at Q3

- ● Inpatient
- ● Discharged

The patient has been enrolled in the trial for **NaN** days

3.1.1 Date follow-up form completed Q3.1.1 is only completed if patient is still an inpatient at Q3

yyyy-mm-dd

28/05/2020 Follow-up

Hydroxychloroquine for COVID-19

3.1.1 What was the date of discharge? Q3.1.1 is only completed if patient has been discharged at Q3

yyyy-mm-dd

3.1 What was the date of death? * Q3.1.1 is only completed if patient has died at Q3

yyyy-mm-dd

3.2 What was the underlying cause of death? *

This can be obtained from the last entry in part 1 of the death certificate

- ◯ COVID-19
- ◯ Other infection
- ◯ Cardiovascular
- ◯ Other

Please give details

4. Did the patient require any form of assisted ventilation (ie, more than just supplementary oxygen)? *

- ◯ Yes
- ◯ No

Please answer the following questions:

4.1 For how many days did the patient require assisted ventilation? *

4.2 What type of ventilation did the patient receive?

	Yes	No	Unknown
CPAP alone	◯	◯	◯
Non-invasive ventilation (eg, BiPAP)	◯	◯	◯
High-flow nasal oxygen (eg, AIRVO)	◯	◯	◯
Mechanical ventilation (intubation/tracheostomy)	◯	◯	◯
ECMO	◯	◯	◯

Hydroxychloroquine for COVID-19

Total number of days the patient received invasive mechanical ventilation (intubation/tracheostomy) (from randomisation until discharge/death/28 days after randomisation)

Complete if invasive mechanical ventilation (intubation/tracheostomy) is Yes

5. Has the participant been documented to have a NEW cardiac arrhythmia at any point since the main randomisation?

- ◯ Yes
- ◯ No
- ◯ Unknown

5. Has the participant been documented to have a NEW cardiac arrhythmia at any point since the main randomisation?

- ◯ Yes
- ◯ No
- ◯ Unknown

5.1 Please select all of the following which apply

If Q5 is answered Yes, you must select at least one option here

- ☐ Atrial flutter or atrial fibrillation
- ☐ Supraventricular tachycardia
- ☐ Ventricular tachycardia (including torsades de pointes)
- ☐ Ventricular fibrillation
- ☐ Atrioventricular block requiring intervention (eg, cardiac pacing)

6. Did the patient require use of renal dialysis or haemofiltration?

- ◯ Yes
- ◯ No

7. Please enter UKOSS case ID if known *

Enter the full UKOSS case ID ie, COR_123

Complete only if patient was pregnant at randomisation

(select if you do not know the UKOSS case ID)

☐ Not known

Cause of death

Cause of death was recorded by the site staff on the Follow-up form. In addition, information about cause of death was obtained from death registration data in England, Wales and Scotland. Where cause of death information was available from both sources, the underlying cause of death from the death registration data was used (in preference to what was recorded on the Follow-up form). In the death registration data, the underlying cause of death is based on the death certificate information completed by the certifying doctor and is recorded using International Classification of Disease 10 codes. These were grouped into relevant categories as described in the Recovery Definition and Derivation of Baseline Characteristics and Outcomes document (see https://www.recoverytrial.net).

Supplementary Exercise 4.19

The full article **Repurposed antiviral drugs for COVID-19 - interim WHO SOLIDARITY trial results** is found here.

MedRxiv (October 15) version

Repurposed antiviral drugs for COVID-19 –interim WHO SOLIDARITY trial results

WHO Solidarity trial consortium*

ABSTRACT

BACKGROUND

WHO expert groups recommended mortality trials in hospitalized COVID-19 of four re-purposed antiviral drugs.

METHODS

Study drugs were Remdesivir, Hydroxychloroquine, Lopinavir (fixed-dose combination with Ritonavir) and Interferon-β1a (mainly subcutaneous; initially with Lopinavir, later not). COVID-19 inpatients were randomized equally between whichever study drugs were locally available and open control (up to 5 options: 4 active and local standard-of-care). The intent-to-treat primary analyses are of in-hospital mortality in the 4 pairwise comparisons of each study drug vs its controls (concurrently allocated the same management without that drug, despite availability). Kaplan-Meier 28-day risks are unstratified; log-rank death rate ratios (RRs) are stratified for age and ventilation at entry.

RESULTS

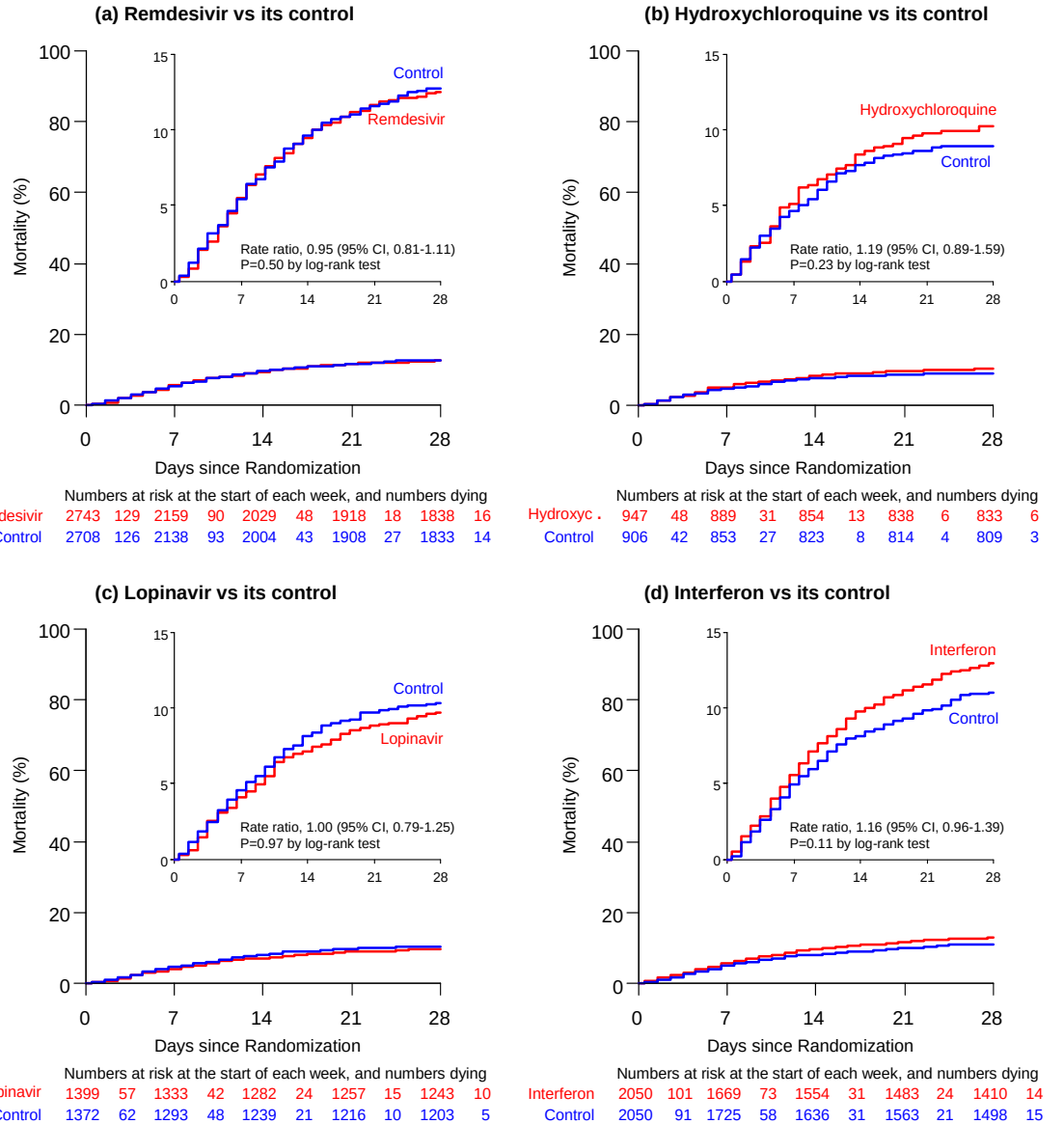
In 405 hospitals in 30 countries 11,266 adults were randomized, with 2750 allocated Remdesivir, 954 Hydroxychloroquine, 1411 Lopinavir, 651 Interferon plus Lopinavir, 1412 only Interferon, and 4088 no study drug. Compliance was 94-96% midway through treatment, with 2-6% crossover. 1253 deaths were reported (at median day 8, IQR 4-14). Kaplan-Meier 28-day mortality was 12% (39% if already ventilated at randomization, 10% otherwise). Death rate ratios (with 95% CIs and numbers dead/randomized, each drug vs its control) were: Remdesivir RR=0.95 (0.81-1.11, p=0.50; 301/2743 active vs 303/2708 control), Hydroxychloroquine RR=1.19 (0.89-1.59, p=0.23; 104/947 vs 84/906), Lopinavir RR=1.00 (0.79-1.25, p=0.97; 148/1399 vs 146/1372) and Interferon RR=1.16 (0.96-1.39, p=0.11; 243/2050 vs 216/2050). No study drug definitely reduced mortality (in unventilated patients or any other subgroup of entry characteristics), initiation of ventilation or hospitalisation duration.

CONCLUSIONS

These Remdesivir, Hydroxychloroquine, Lopinavir and Interferon regimens appeared to have little or no effect on hospitalized COVID-19, as indicated by overall mortality, initiation of ventilation and duration of hospital stay. The mortality findings contain most of the randomized evidence on Remdesivir and Interferon, and are consistent with meta-analyses of mortality in all major trials. (Funding: WHO. Registration: ISRCTN83971151, NCT04315948)

Figure 2. Effects of (a) Remdesivir, (b) Hydroxychloroquine, (c) Lopinavir and (d) Interferon on 28-day mortality

Kaplan-Meier graphs of in-hospital mortality. The inset shows the same data on an expanded y-axis.



INTRODUCTION

A WHO COVID-19 research forum in February 2020 recommended evaluation of treatments in large randomized trials, and other WHO expert groups identified 4 re-purposed anti-viral drugs that might have at least a moderate effect on mortality: Remdesivir, Hydroxychloroquine, Lopinavir, and Interferon-β1a.

In March 2020, WHO began a large, simple, multi-country, open-label randomized trial among hospital inpatients of the effects of these 4 drugs on in-hospital mortality. The trial was adaptive; unpromising drugs could be dropped and others added. Hydroxychloroquine and Lopinavir were eventually dropped, but others, such as monoclonal antibodies, will be added. We report interim mortality results for the original 4 drugs.

METHODS

The protocol was designed to involve hundreds of potentially over-stressed hospitals in dozens of countries. Hence, no form-filling was required, and trial procedures were minimal but rigorous. Online randomization of consented patients (via a cloud-based GCP-compliant clinical data management system) took just a few minutes, as did online reporting of death in hospital or discharge alive (plus brief details of respiratory support in hospital and use of study drugs and certain non-study drugs). No other reporting was required unless doctors suspected an unexpected serious adverse reaction (SUSAR). National and global monitors resolved queries and checked progress and data completeness. Eligible patients were age $\geq$ 18 years, hospitalized with a diagnosis of COVID-19, not known to have received any study drug, without anticipated transfer elsewhere within 72 hours, and, in the physician's view, with no contra-indication to any study drug. Participants were randomized in equal proportions between control and whichever other study drugs were locally available (up to 5 options: these drugs, and local standard-of-care). Placebos were not used. Study drugs were Remdesivir, Hydroxychloroquine, Lopinavir-Ritonavir and Interferon (given with Lopinavir, until July 4). Hydroxychloroquine and Lopinavir were discontinued for futility on June 18 and July 4, 2020, respectively; Interferon is ceasing on October 16.

Daily doses were those already used for other diseases, but to maximize any efficacy without undue cardiac risk Hydroxychloroquine dosage was based on that for amoebic liver abscess, rather than the lower dosage for malaria. (Hydroxychloroquine slightly prolongs QT, and unduly high or rapid dosage might cause arrhythmias or hypotension.) All treatments were stopped at discharge; otherwise, regimens were:

- Remdesivir (intravenous): Day 0, 200mg; days 1-9, 100mg.
- Hydroxychloroquine (oral): Hour 0, four tablets; Hour 6, four tablets; Hour 12, begin two tablets twice daily for 10 days. Each tablet contained 200mg Hydroxychloroquine sulphate (155mg base/tablet; a little-used alternative involved 155mg chloroquine base/tablet).
- Lopinavir (oral): Two tablets twice daily for 14 days. Each tablet contained 200mg Lopinavir (plus 50mg Ritonavir, to slow hepatic clearance of Lopinavir). Other formulations were not provided, so ventilated patients received no study Lopinavir while unable to swallow.
- Interferon (mainly subcutaneous): Three doses over six days of 44μg subcutaneous Interferon-β1a; where intravenous interferon was available, patients on high-flow oxygen, ventilators or ECMO were instead to be given 10μg intravenously once daily for six days.

ENDPOINTS

The protocol-specified primary objective was to assess effects on in-hospital mortality (ie, mortality during the original episode of hospitalization; follow-up ceased at discharge) not only in all patients but also in those with moderate COVID and in those with severe COVID (subsequently defined as ventilated when randomized). The protocol-specified secondary outcomes were initiation of ventilation and hospitalization duration. Although no placebos were used, appropriate analyses of these non-fatal outcomes can still be reliably informative. The CATCO add-on study in Canada and the Discovery add-on study in Europe (mostly France) recorded additional outcomes that will be reported elsewhere.

SAMPLE SIZE

protocol stated "The larger the number entered the more accurate the results will be, but numbers entered will depend on how the epidemic develops... it may be possible to enter several thousand hospitalised patients with relatively mild disease and a few thousand with severe disease, but realistic, appropriate sample sizes could not be estimated at the start of the trial." The Executive Group, blind to any findings, decided the timing of release of interim results.

STATISTICAL ANALYSES

The four main sets of analyses involve the evenly randomized pairwise comparisons of each study drug vs its controls. The controls for those randomly allocated one particular drug were those patients who could by chance have been randomly allocated that drug (at that moment, in that hospital), but instead got allocated standard of care. If, for a particular study entrant, more than one study drug was available, allocation to standard of care would put that patient into the control group for each of them. Hence, there is partial overlap between the four control groups. Each comparison between a study drug and its controls, however, is evenly randomized (50/50) and unbiased, as both groups are affected equally by any differences between countries or

hospitals and by any time trends in patient characteristics or standard of care.

All analyses relate mortality to allocated treatment (ie, they are intent-to-treat analyses). The overall mortality analyses were of all randomised patients (drug vs its control), and the only protocol-specified subgroup analyses are those considering separately patients with moderate and with severe COVID (ie, already ventilated; the type of ventilation was not recorded at study entry.) Unstratified Kaplan-Meier methods plot 28-day risk. Death rate ratios (RRs) and p-values are from log-rank analyses, stratified for 3x2=6 strata of age and ventilation at entry. If the stratified log-rank Observed minus Expected number of deaths is O-E with variance V, logeRR is calculated as (O-E)/V with variance 1/V and a The few currently uncertain death times were taken as day 7. Analyses censored patients with outcome not yet reported at day 0, and censored the few inter-hospital transfers at transfer. They did not censor patients discharged alive, as analyses were of mortality during the initial hospitalisation. Forest plots (with 95% CIs only for overall results, otherwise 99% CIs) and chi-squared statistics (sum of [O-E]2/V, with no p-value given) help interpret any apparent heterogeneity of treatment RRs between subgroups. Analyses used SASv9.4 and Rv4.02.

The Discussion includes meta-analyses of the major trial results, based on the inverse-variance-weighted average of b=logeRR from each stratum of each trial, using odds ratios where hazard or death rate ratios were unavailable. (This weighted average is derived from the sums of [O-E] and of V over strata.) In general, the more deaths in a stratum the larger V is and, correspondingly, the smaller is the variance of logeRR, so the more weight that stratum gets. The variance attributed to the result in each stratum and to the overall weighted average reflects only the play of chance at randomization. Homogeneity of different RRs is not needed for this weighted average to be informative.

Figure 3. Rate ratios of any death, stratified by age and respiratory support at entry, (a) Remdesivir, (b) Hydroxychloroquine, (c) Lopinavir, (d) Interferon, each vs its control

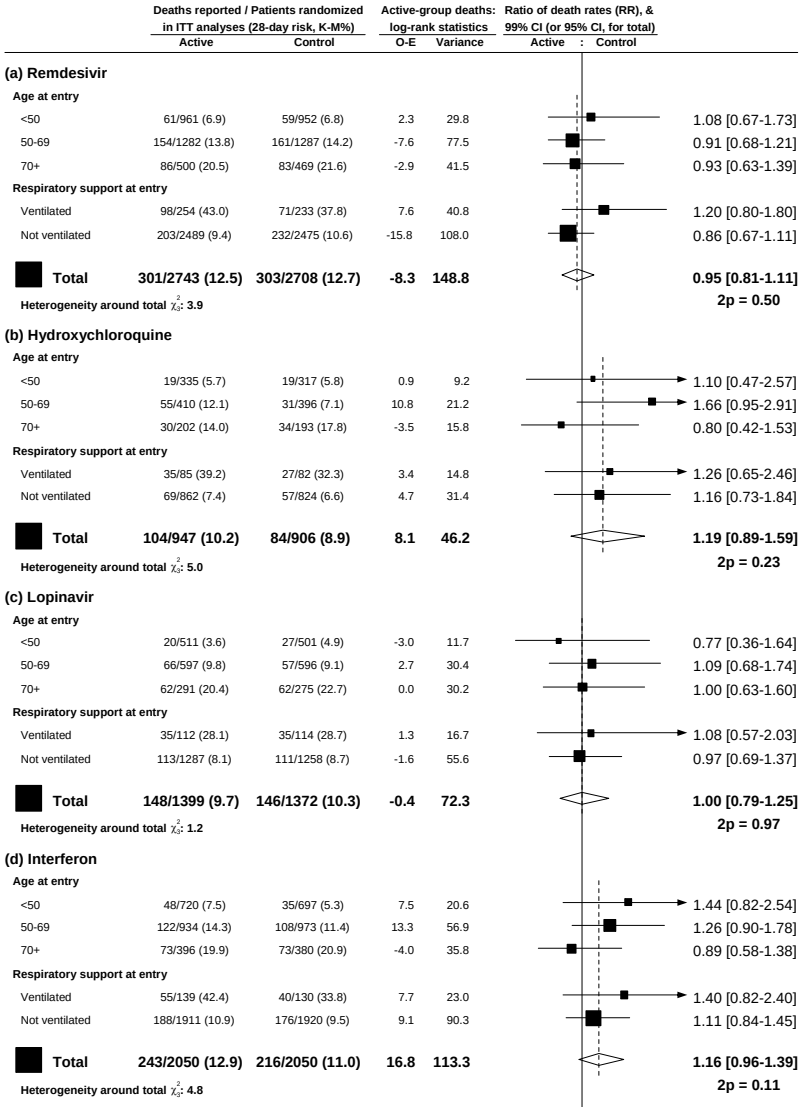
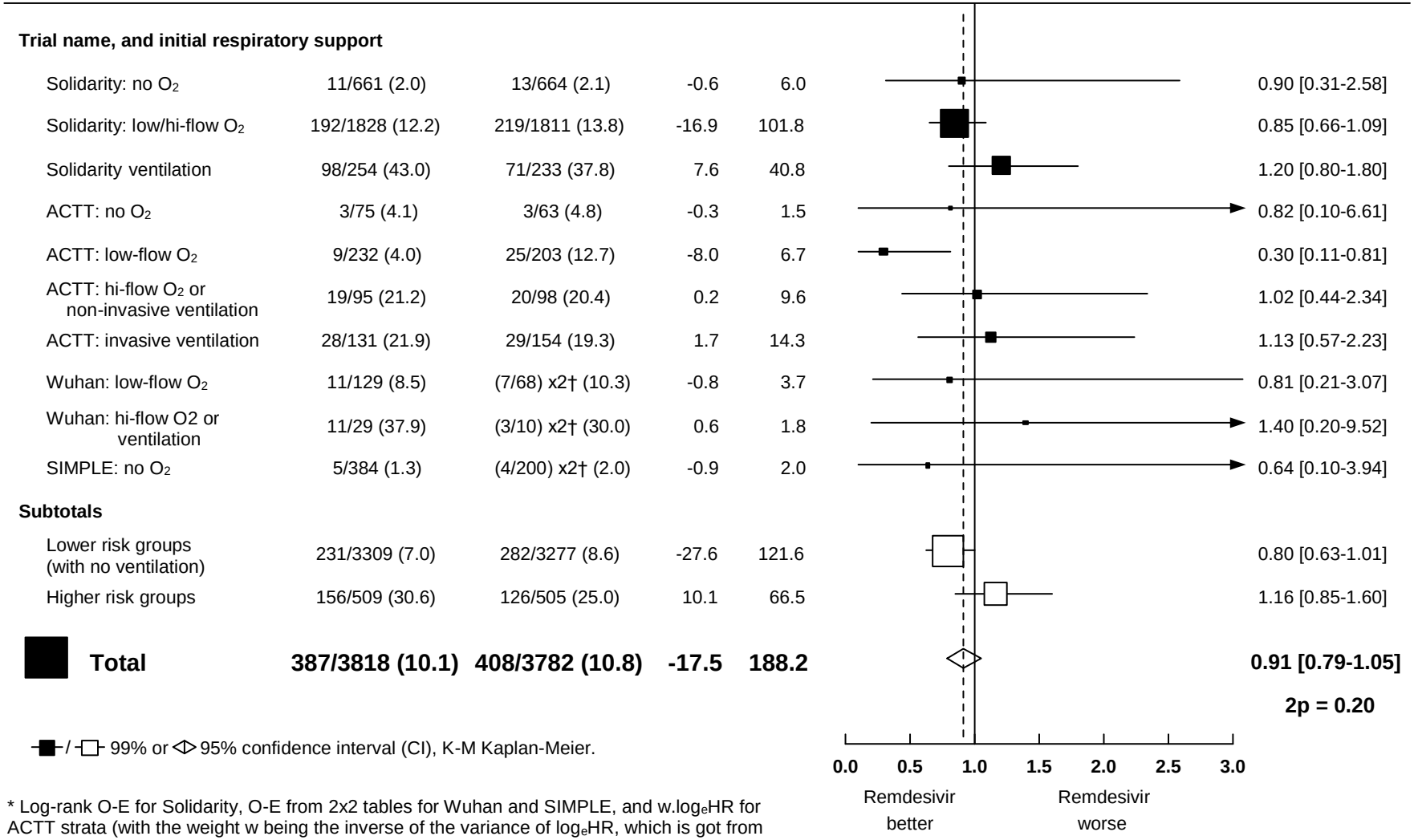


Figure 4. Remdesivir vs control – Meta-analysis of mortality in trials of random allocation of hospitalised COVID-19 patients to Remdesivir or the same treatment without it

	Deaths reported / Patients randomized in ITT analyses (28-day risk, K-M%)		Remdesivir deaths: Observed-Expected		Ratio of death rates (RR), & 99% CI (or 95% CI, for total)
	Remdesivir	Control	(O-E)*	Var (O-E)	Remdesivir : Control
Trial name, and initial respiratory support					
Solidarity: no O_2	11/661 (2.0)	13/664 (2.1)	-0.6	6.0	0.90 [0.31-2.58]
Solidarity: low/hi-flow O_2	192/1828 (12.2)	219/1811 (13.8)	-16.9	101.8	0.85 [0.66-1.09]
Solidarity ventilation	98/254 (43.0)	71/233 (37.8)	7.6	40.8	1.20 [0.80-1.80]
ACTT: no O_2	3/75 (4.1)	3/63 (4.8)	-0.3	1.5	0.82 [0.10-6.61]
ACTT: low-flow O_2	9/232 (4.0)	25/203 (12.7)	-8.0	6.7	0.30 [0.11-0.81]
ACTT: hi-flow O_2 or non-invasive ventilation	19/95 (21.2)	20/98 (20.4)	0.2	9.6	1.02 [0.44-2.34]
ACTT: invasive ventilation	28/131 (21.9)	29/154 (19.3)	1.7	14.3	1.13 [0.57-2.23]
Wuhan: low-flow O_2	11/129 (8.5)	(7/68) x2† (10.3)	-0.8	3.7	0.81 [0.21-3.07]
Wuhan: hi-flow O2 or ventilation	11/29 (37.9)	(3/10) x2† (30.0)	0.6	1.8	1.40 [0.20-9.52]
SIMPLE: no O_2	5/384 (1.3)	(4/200) x2† (2.0)	-0.9	2.0	0.64 [0.10-3.94]
Subtotals					
Lower risk groups (with no ventilation)	231/3309 (7.0)	282/3277 (8.6)	-27.6	121.6	0.80 [0.63-1.01]
Higher risk groups	156/509 (30.6)	126/505 (25.0)	10.1	66.5	1.16 [0.85-1.60]
Total	**387/3818 (10.1)**	**408/3782 (10.8)**	**-17.5**	**188.2**	**0.91 [0.79-1.05]**
					2p = 0.20



■/□ 99% or ◇ 95% confidence interval (CI), K-M Kaplan-Meier.

* Log-rank O-E for Solidarity, O-E from 2x2 tables for Wuhan and SIMPLE, and w.log$_e$HR for ACTT strata (with the weight w being the inverse of the variance of log$_e$HR, which is got from the HR's CI). RR is got by taking log$_e$RR to be (O-E)/V with Normal variance 1/V. Subtotals or totals of (O-E) and of V yield inverse-variance-weighted averages of the log$_e$RR values.

† For balance, controls in the 2:1 studies count twice in the control totals and subtotals.

Supplementary Exercise 4.20

The full article **Survival of SARS-CoV-2 and influenza virus on the human skin: Importance of hand hygiene in COVID-19** is found here.

Ryohei Hirose,[1,2*] Hiroshi Ikegaya,[3] Yuji Naito,[2] Naoto Watanabe, [1,2] Takuma Yoshida, [1,2] Risa Bandou,[1,3] Tomo Daidoji,[1] Yoshito Itoh,[2] Takaaki Nakaya[1]

[1] Department of Infectious Diseases, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, 465 Kajii-cho, Kawaramachi-Hirokoji, Kamigyo-ku, Kyoto 602-8566, Japan.

[2] Department of Molecular Gastroenterology and Hepatology, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, 465 Kajii-cho, Kawaramachi-Hirokoji, Kamigyo-ku, Kyoto 602-8566, Japan.

[3]Department of Forensics Medicine, Graduate School of Medical Science, Kyoto Prefectural University of Medicine, 465 Kajii-cho, Kawaramachi-Hirokoji, Kamigyo-ku, Kyoto 602-8566, Japan.

Summary

The survival time of SARS-CoV-2 on the human skin was approximately 9 h, significantly longer than that of IAV (approximately 1.8 h). The longer survival of SARS-CoV-2 on the skin increases contact-transmission risk; however, hand hygiene can reduce this risk.

Figure 1. Outline of the pathogen stability evaluation model and its reproducibility. The pathogen stability evaluation model was constructed using human skin collected from forensic autopsy specimens (A). To evaluate the reproducibility of the model, influenza A virus (IAV) was applied to the six model skin samples and to the hand skin of six subjects (amount of virus: 1.0×10^5 FFU), and the titer of the remaining viruses on the skin was measured. The 95% confidence interval (red bar) of the viable virus titer on the model skin at each elapsed time was within the 95% confidence interval (blue bar) of the viable virus titer on the skin of live individuals (B).

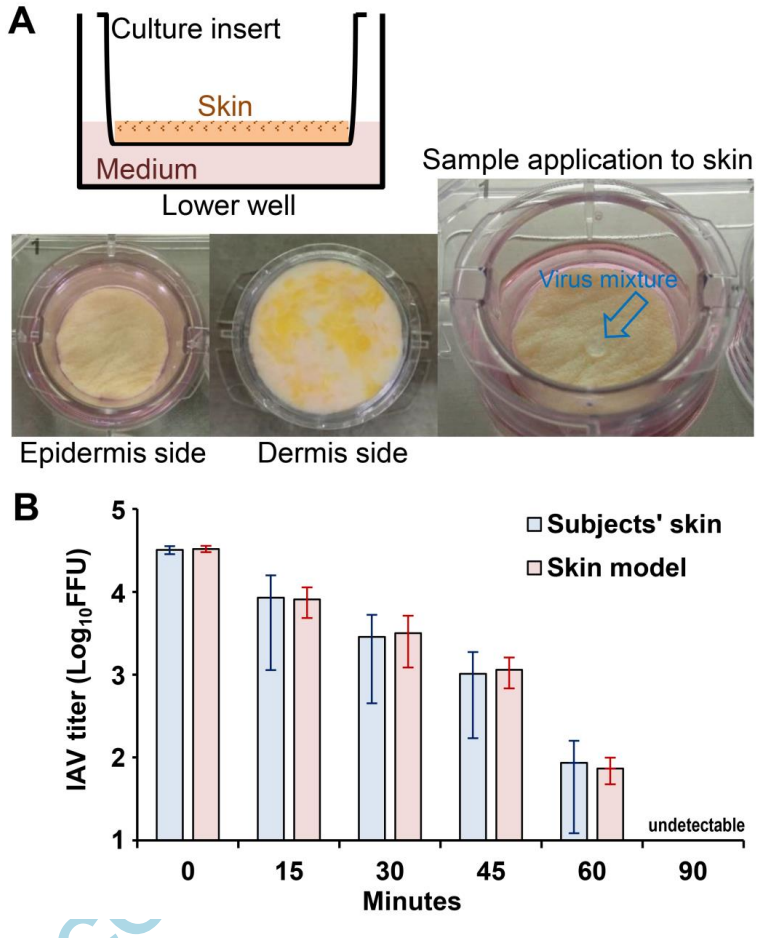


Figure 2. (A–F) Fluctuations in the titer of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and influenza A virus (IAV) surviving on the surface of stainless steel (A), borosilicate glass (B), polystyrene (C), and three skin samples [HS1 (D), HS2 (E), and HS3 (F)]. SARS-CoV-2/IAV was mixed with Dulbecco's modified Eagle's medium (DMEM) or mucus and applied in 5-μL aliquots to each surface (amount of virus: 1.0×10^5 FFU or 1.0×10^5 $TCID_{50}$, respectively). Each surface was incubated in a constant environment (temperature: 25 °C, humidity: 45–55%) for 0–120 h. The remaining viruses on the surface were then recovered in 1 ml of culture medium and titrated. For each measurement, three independent experiments were performed, and the results are expressed as the mean ± standard error of the mean. Bars referring to the data below the detection limit were omitted. See Supplementary Figure S1 and S2 for raw data.

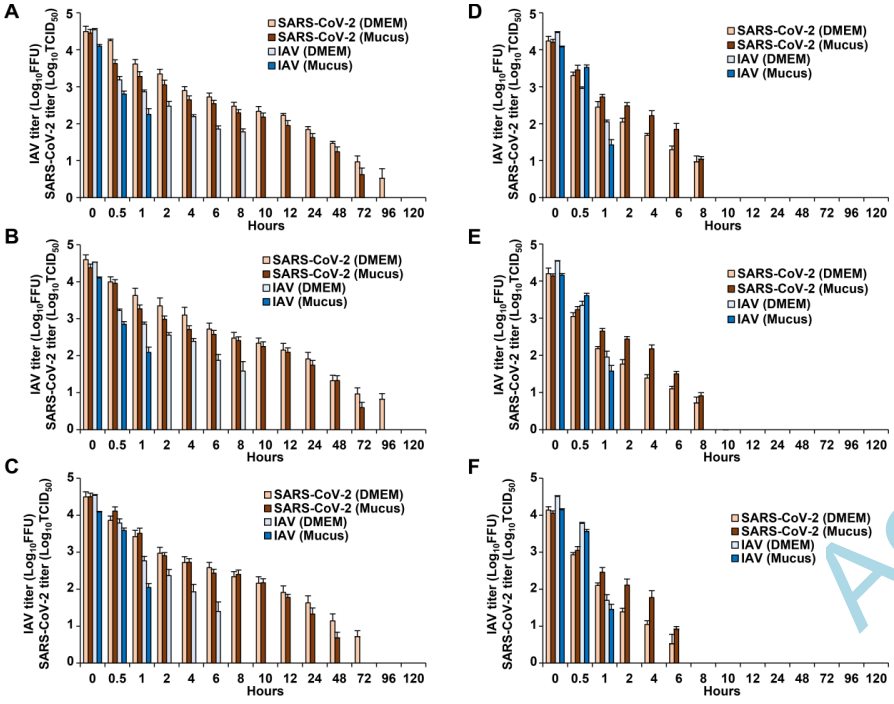


Figure 3. Evaluation of the disinfection effectiveness of 80% (w/w) ethanol against SARS-CoV-2 (upper panel) and IAV (lower panel) on human skin. Thirty minutes after the mixture of the DMEM/mucus and SARS-CoV-2/IAV was applied to each skin surface (HS1/HS2/HS3), 80% ethanol was further applied to the skin surfaces for 15 s, followed by disinfectant inactivation via dilution with culture medium. The surviving viruses on the skin surfaces were then titrated. For comparison, the surviving viruses on the skin surfaces in the absence of ethanol were also titrated over time. SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; IAV, influenza A virus; DMEM, Dulbecco's modified Eagle's medium. For each measurement, three independent experiments were performed, and the results are expressed as mean ± standard error values.

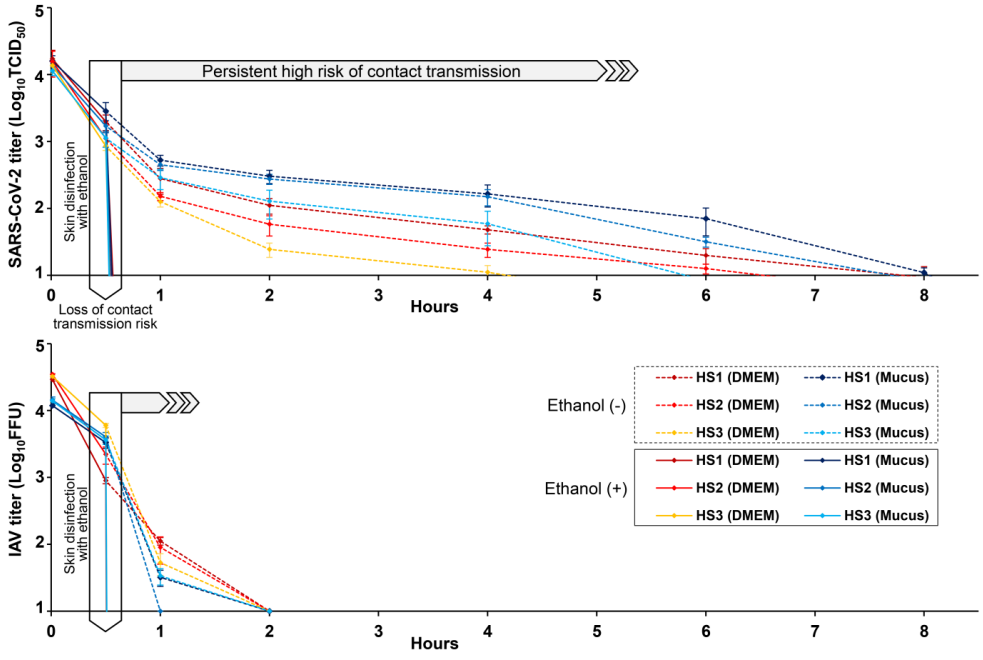


Table 1. Survival time and half-life time of viruses on each surface.

	Survival time[1], hour, median (95% CI)				Half-life time[2], hour, median (95% CI)			
	IAV (DMEM)	**SARS-CoV-2 (DMEM)**	**IAV (Mucus)**	**SARS-CoV-2 (Mucus)**	**IAV (DMEM)**	**SARS-CoV-2 (DMEM)**	**IAV (Mucus)**	**SARS-CoV-2 (Mucus)**
Stainless steel	11.56 (10.11-13.22)	84.29 (54.01-119.56)	1.73 (1.57-1.91)	64.51 (52.35-77.73)	6.78 (5.84-7.97)	32.62 (16.80-56.68)	0.86 (0.76-0.98)	25.53 (18.45-34.24)
Borosilicate glass	10.61 (9.18-12.27)	85.74 (56.27-119.80)	1.73 (1.58-1.88)	61.23 (49.03-74.44)	6.13 (5.22-7.29)	33.24 (17.59-56.49)	0.85 (0.76-0.96)	23.63 (17.16-31.86)
Polystyrene	6.07 (5.05-7.27)	58.07 (37.76-81.95)	1.96 (1.76-2.18)	35.92 (29.58-42.67)	3.04 (2.40-3.87)	22.58 (11.64-41.24)	0.91 (0.80-1.04)	13.17 (10.26-17.35)
Human skin (HS total)	1.82 (1.65-2.00)	9.04 (7.96-10.22)	1.69 (1.57-1.81)	11.09 (10.22-12.00)	0.80 (0.72-0.90)	3.53 (3.02-4.16)	0.77 (0.71-0.84)	4.16 (3.79-4.58)
Human skin (HS1)	1.81 (1.64-2.00)	10.93 (8.95-13.10)	1.66 (1.47-1.88)	12.24 (10.64-13.94)	0.82 (0.73-0.93)	4.13 (3.29-5.28)	0.77 (0.66-0.89)	4.47 (3.83-5.26)
Human skin (HS2)	1.79 (1.50-2.13)	9.45 (7.72-11.38)	1.71 (1.51-1.94)	12.2 (11.10-13.34)	0.78 (0.64-0.98)	3.75 (2.93-4.86)	0.78 (0.67-0.91)	4.51 (4.06-5.03)
Human skin (HS3)	1.86 (1.50-2.27)	6.14 (4.91-7.53)	1.69 (1.49-1.91)	8.13 (6.85-9.51)	0.79 (0.63-1.04)	2.36 (1.73-3.21)	0.77 (0.67-0.90)	3.13 (2.56-3.86)

The elapsed time was defined as an explanatory variable (X-axis), and the log virus titer of IAV or SARS-CoV-2 was defined as an explained variable (Y-axis). A linear regression analysis with logarithmic link function was performed for each virus to create a curve of regression (see also Supplementary Figure S3).

[1]The measurement limits of the titers of IAV and SARS-CoV-2 were 10^1 FFU and $10^{0.5}$ $TCID_{50}$, respectively; therefore, the survival times of IAV and SARS-CoV-2 were defined as the X values when the Y values of the regression curves were 1.0 and 0.5, respectively.

[2]The half-life time of each log virus titer was calculated from the slope of each regression line.

Supp Exerc. 4.21: A Monoclonal Antibody for Malaria Prevention

PART 1 (2021 REPORT) The full version of the first article, in 2021, on the proof of principle, on safety, the initial side-effect profile, and pharmacokinetics in healthy adults who had never had malaria, can be found here. To assess the protective efficacy of CIS43LS, some (15) participants underwent controlled human malaria infection in which they were exposed to mosquitoes carrying P. falciparum sporozoites 4 to 36 weeks after administration of CIS43LS.

Controlled Human Malaria Infection

Participants were exposed to bites on the forearm from Anopheles stephensi mosquitoes infected with P. falciparum (3D7 strain). The mosquitoes met standard infectivity criteria as previously described. Outpatient monitoring was performed by means of two telephone calls in the first 7 days after infection challenge, followed by clinic visits on days 7 through 18 and on day 21 to assess for parasitemia with standard polymerase-chain- reaction (PCR) methods. Parasitemia (i.e., malaria infection) was defined as a single positive PCR result. Participants were considered protected if they remained negative for parasitemia through day 21 after infection. Directly observed treatment with 1 g of atovaquone and 400 mg of proguanil hydrochloride was administered for 3 consecutive days, beginning at the time parasitemia was confirmed or on day 21 if the participant had not already been treated.

(The target sample size was determined on the basis of the probability of observing serious adverse events.) The efficacy analysis included all enrolled participants who received CIS43LS and underwent controlled human malaria infection. The primary efficacy analysis was performed with the use of a Barnard test to assess the percentage of participants who had malaria infection. The secondary efficacy analysis was performed with the use of a log-rank test to compare the time to parasitemia among participants who received CIS43LS with that among control participants. The salivary gland scores for the mosquitoes used in controlled infections are reported, along with the median values and interquartile ranges.

Controlled human malaria infection was administered to 15 participants (9 who had received CIS43LS and 6 control participants) on October 20, 2020. The 21-day monitoring for parasitemia concluded on November 10, 2020. One participant who received 40 mg per kilogram intravenously in Part B did not undergo controlled infection because of a concomitant illness. Participants in Part B were followed through March 2021. Maximum enrollment was not met in Part B because of Covid-19–related restrictions.

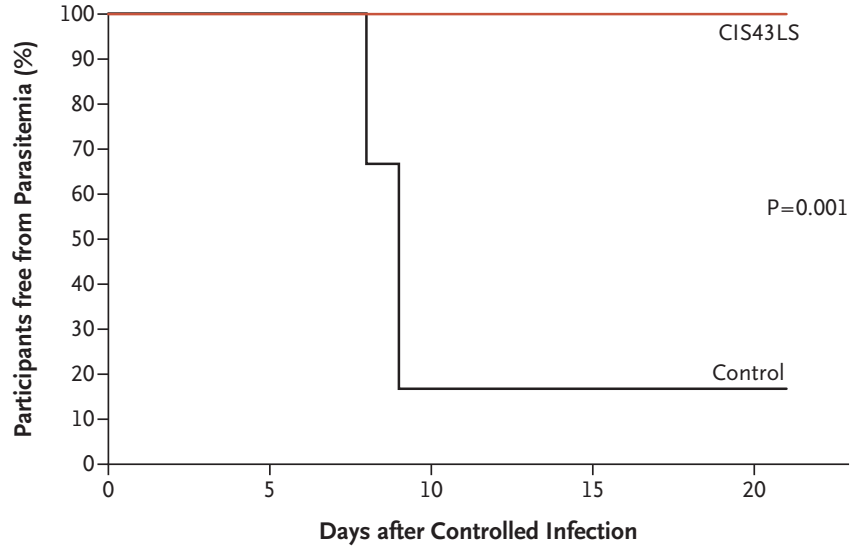


Figure 4. Parasitemia after Controlled Human Malaria Infection.
A Kaplan–Meier analysis shows the time to parasitemia as measured by polymerase-chain-reaction analysis. A log-rank test comparing parasitemia among the nine participants who received CIS43LS with that among the six control participants yielded a P value of 0.001.

None of the 9 participants who underwent controlled human malaria infection and had received CIS43LS had parasitemia through day 21, whereas parasitemia developed in 5 of 6 control participants on days 8 or 9 after infection ($P = 0.001$ by two-sided Barnard test), a finding consistent with historical data for control participants who underwent infection through this model (Fig. 4).

All participants who underwent controlled infection met pre-specified malaria exposure criteria at the time of the challenge, which consisted of five qualifying bites from mosquitoes with a salivary gland score of 2 or greater (scores range from 0 to 4, with higher scores indicating more microscopically observed sporozoites). The median salivary gland score was 3.2 (interquartile range, 2.6 to 3.2) in mosquitoes that bit participants who had received CIS43LS and 3.1 (interquartile range, 3.0 to 3.4) in mosquitoes that bit control participants (Table S3). At the time of controlled infection, the serum concentrations of CIS43LS ranged from approximately 50 to 500 μg per milliliter among the 9 participants who had received CIS43LS. Two participants who underwent controlled infection up to 36 weeks after administration of CIS43LS had serum concentrations of approximately 50 μg per milliliter at the time of infection.

PART 2 (2022 REPORT)

The full version of the 2022 report addressed the safety and efficacy of a single intravenous infusion of CIS43LS against P. falciparum infection in healthy adults in Mali over a 6-month malaria season. It can be found here. In Part A, safety was assessed at three escalating dose levels. In Part B, participants were randomly assigned (in a 1:1:1 ratio) to receive 10 mg of CIS43LS per kilogram of body weight, 40 mg of CIS43LS per kilogram, or placebo. The primary efficacy end point, assessed in a time-to-event analysis, was the first P.falciparum infection detected on blood-smear examination, which was performed at least every 2 weeks for 24 weeks.

Randomized, Placebo-Controlled Trial (Part B)

In Part B, 330 participants were randomly assigned (in a 1:1:1 ratio) by block randomization to receive 10 mg of CIS43LS per kilogram, 40 mg of CIS43LS per kilogram, or placebo (110 participants in each group) by intravenous infusion. Trial participants and trial team members were unaware of the trial-group assignments. Only the pharmacists preparing the trial agents were aware of such assignments. The pharmacists prepared CIS43LS and the normal saline placebo (both colorless) using identical infusion bags that contained the same volume. Participants received a single infusion of CIS43LS or placebo (day 0) and were followed at trial visits 1, 3, 7, 14, 21, and 28 days later and then once every 2 weeks thereafter through 24 weeks. Primary trial assessments included physical examination and blood collection for the detection of P. falciparum by microscopic examination of thick blood smears. Blood smears were analyzed by two independent readers who were unaware of the trial-group assignments. A third reader examined blood smears when discrepancies occurred. A positive blood smear was defined as two independent readers both reporting the presence of any P. falciparum asexual parasites after counting 2500 leukocytes or examining 200 high-power fields. The competency of blood-smear readers is regularly assessed at the Mali Research and Training Center laboratory, which is certified by the College of American Pathologists.

In Parts A and B, all the participants received a standard treatment course of artemether-lumefantrine 7 to 21 days before administration of CIS43LS or placebo to clear possible P. falciparum blood-stage infection. The administration of all doses of artemether-lumefantrine was directly observed by trial staff. For the remainder of the trial, asymptomatic P. falciparum infections were not treated, in accordance with national guidelines in Mali. All the participants in whom symptomatic malaria developed during the trial were provided standard treatment.

The pre-specified primary efficacy analysis used the modified intention-to-treat data set and was based on the time to the first P. falciparum infection. P values that are reported for the primary efficacy end point were based on the log-rank test comparing each CIS43LS group with the placebo group. Protective efficacy was estimated by the hazard ratio from the Cox proportional-hazards model that accounted for interval censoring. Time-to-event efficacy was calculated as efficacy (%)=(1-HR)×100, in which HR is the hazard ratio for infection between trial groups. Detailed statistical methods are provided in the Supplementary Appendix.

(Approx.) data JH was able to reconstruct from Figure 2

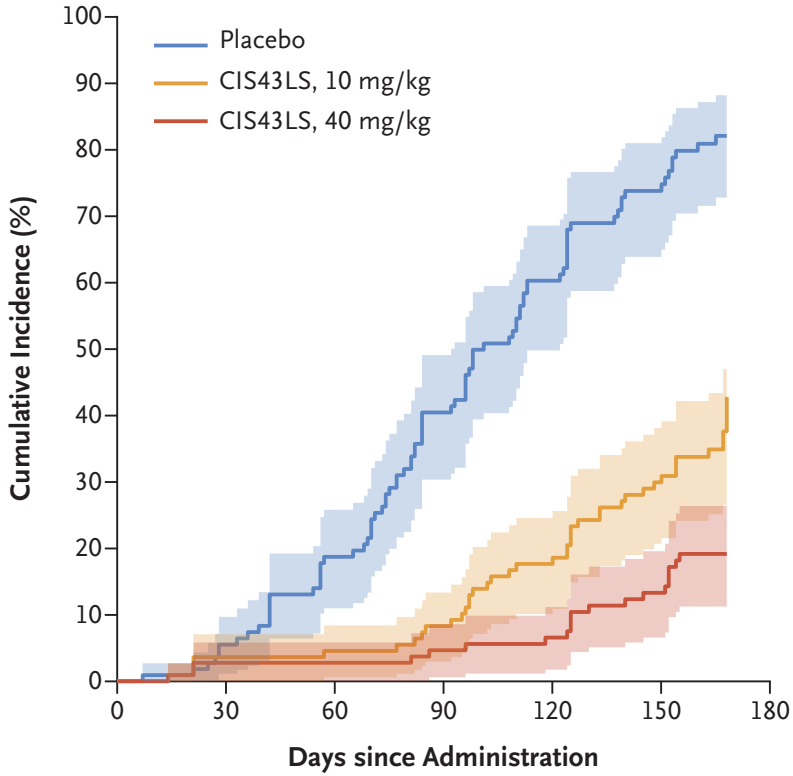


Figure 2. Kaplan–Meier Plot of Efficacy against *P. falciparum* Infection.

Shown is the cumulative incidence of *P. falciparum* blood-stage infection during a 6-month malaria season (irrespective of symptoms being present) after a single intravenous infusion of 10 mg of CIS43LS per kilogram, 40 mg of CIS43LS per kilogram, or placebo. *P. falciparum* infections were detected by microscopic examination of thick blood smears collected during scheduled trial visits and unscheduled illness visits. Blood smears were collected before the administration of CIS43LS or placebo on day 0 and then on days 3, 7, 14, 21, and 28 and every 2 weeks thereafter for a total of 24 weeks. Only blood smears collected between weeks 1 and 24 were included in the efficacy analysis. Shaded areas indicate the 95% confidence intervals.

SAFETY

In Part B, solicited local and systemic adverse events within 7 days after administration of CIS43LS or placebo were all mild to moderate in severity (Table 2) and, apart from headache, were similar in frequency across trial groups. The risk of moderate headache was 3.3 times as high with 40 mg of CIS43LS per kilogram as with placebo (unadjusted 95% confidence interval [CI], 1.1 to 9.7). All solicited adverse events resolved. From the time that CIS43LS or placebo was administered through the end of the 24-week trial period, there were 1235 unsolicited adverse events: 342 grade 1 (27.7%), 880 grade 2 (71.3%), 12 grade 3 (1.0%), and 1 grade 5 (0.1%) (Table S4). There were 4 serious adverse events (Table S5), all considered by investigators to be unrelated to the trial in blinded investigations. [...]

EFFICACY

Among the 330 participants included in the modified intention-to-treat data set, P. falciparum infections detected on blood-smear examination with an onset between weeks 1 and 24 after administration of the active drug or placebo occurred in 39 participants (35.5%) who received 10 mg of CIS43LS per kilogram, 20 (18.2%) who received 40 mg of CIS43LS per kilogram, and 86 (78.2%) who received placebo.

In the primary efficacy analysis that was based on the time to the first P. falciparum infection over the 24-week trial period, the efficacy [(1-HR)×100] of 40 mg of CIS43LS per kilogram as compared with placebo was 88.2% (adjusted 95% CI, 79.3 to 93.3; P<0.001), and the efficacy of 10 mg of CIS43LS per kilogram as compared with placebo was 75.0% (adjusted 95% CI, 61.0 to 84.0; P < 0.001) (Fig. 2). The median P. falciparum parasitemia at the first detected infection after administration of CIS43LS or placebo was similar across trial groups (220 parasites per microliter among those who received 10 mg per kilogram, 160 parasites per microliter among those who received 40 mg per kilogram, and 240 parasites per microliter among those who received placebo).

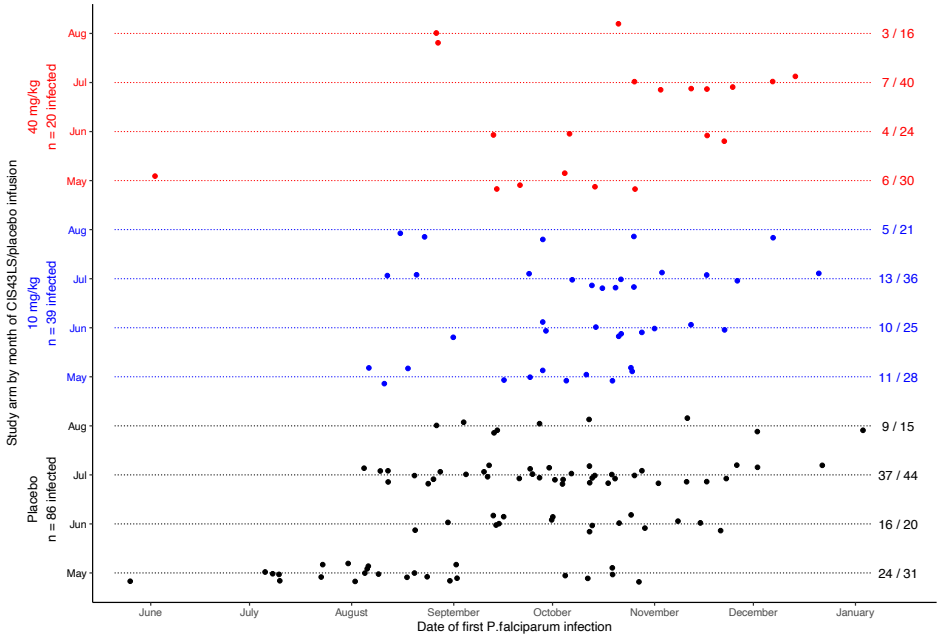
In the secondary efficacy analysis that was based on the Kaplan-Meier estimate of the proportion of participants infected with P. falciparum over the 24-week trial period, the efficacy [(1 - relative risk) × 100] of 40 mg of CIS43LS per kilogram as compared with placebo was 76.7% (adjusted 95% CI, 52.8 to 86.7; P<0.001), and the efficacy of 10 mg of CIS43LS per kilogram as compared with placebo was 54.2% (adjusted 95% CI, 31.1 to 67.6; P<0.001). A **post hoc analysis**, the details of which are provided in the Supplementary Appendix, showed that time-to-infection efficacy of CIS43LS at 12 weeks of follow-up as compared with placebo was 92.3% (unadjusted 95% CI, 78.4 to 97.2) for 40 mg per kilogram and 84.5% (unadjusted 95% CI, 67.1 to 92.7) for 10 mg per kilogram.

Q's (using individual-level data, reconstructed from Fig 2.)

1. How many patients were lost to follow-up before day 168?
2. Contrast the HR-based and K-M-based efficacy values.
3. Why the greater difference between the two measures of the 10 mg/Kg efficacy?
4. For the 10 mg/Kg dose vs. placebo, replicate the implied HR [of (100-75.0)/100 = 0.25]. Compute separate HRs for the 1st and last 12 weeks of follow-up. Comment.
5. Again, for the 10 mg/Kg vs. placebo contrast, and *assuming* a constant-over-the-24-weeks HR, use the (post-fit) `survfit(aready.fitted.model, newdata= ..`) option in `R coxph` to fit a 24-week risk (cumulative incidence) for each of the 2 contrasted arms, and compute the risk ratio. Why is it so different from the 75% Risk Ratio in the primary efficacy analysis? [also, to check the fits, superimpose the PH-fitted cum. incidence curves on the K-M cum. incidence curves.]

Figure S1. The timing of first *P. falciparum* infections by study arm and month of study agent infusion in the efficacy study (Part B) with respect to calendar time.

Shown is the distribution of first *P. falciparum* infections with respect to calendar time across study arms stratified by the month of study agent infusion for each participant who became infected during the study period. In the efficacy study, 330 adults were randomized 1:1:1 and received 10 mg/kg or 40 mg/kg of CIS43LS, or placebo between May 5 and August 6, 2021. Artemether-lumefantrine was given to all participants 7 to 21 days before study agent administration to clear any possible *P. falciparum* blood-stage infection. Among the 330 participants, *P. falciparum* infections detected by blood smear with an onset between weeks 1 and 24 after study agent administration occurred in 86 (78.2%) participants in the placebo group (black dots), 39 (35.5%) participants in the CIS43LS 10 mg/kg group (blue dots), and 20 (18.2%) participants in the CIS43LS 40 mg/kg group (red dots). The ratio of participants infected per participants infused each month is shown on the right-hand side of the figure.



Supplementary Exercise 4.22:

The survival time of chocolates on hospital wards: covert observational study

On Thu, 10 Dec 2020, James Hanley wrote:

Dear Dr Gajendragadkar

Any chance you would still have the dataset from your Christmas 2013 BMJ article to share with me for teaching?. It would brighten up the class I teach on survival analysis.

Best

James Hanley
(I had an article in the 2013 BMJ, (on the longevity of the Titanic survivors) but it never got any traction. I just saw yours mentioned in the 2016 review)

—-

Fri 2020-12-11

Dear James,

Delighted to share (attached as an SPSS file). Just a point to note - the overall median survival time was calculated by only including those chocolates that were eaten, i.e. - we excluded leftover chocolates when calculating survival time. This was a conscious decision after much discussion about whether or not chocolates left-over at the end were informative/non-informatively censored.

I think you meant 2003 Christmas BMJ for your article? Catchy title - times were very different 'pre-Twitter'.

Best, Parag

Supplementary Exercise 4.23:

Effect of oral antimicrobial prophylaxis on surgical site infection after elective colorectal surgery: multicentre, randomised, double blind, placebo controlled trial.

OBJECTIVE: To investigate whether oral antimicrobial prophylaxis as an adjunct to intravenous antibiotic prophylaxis reduces surgical site infections after elective colorectal surgery.

DESIGN: Multicentre, randomised, double blind, placebo controlled trial.

SETTING: 11 university and non-university hospitals in France between 25 May 2016 and 8 August 2019.

PARTICIPANTS: 926 adults scheduled for elective colorectal surgery.

INTERVENTION Patients were randomised to receive either a single 1 g dose of ornidazole (n=463) or placebo (n=463) orally 12 hours before surgery, in addition to intravenous antimicrobial prophylaxis before surgical incision.

MAIN OUTCOME MEASURES The primary outcome was the proportion of patients with surgical site infection within 30 days after surgery. Secondary outcomes included individual types of surgical site infections and major postoperative complications (Clavien-Dindo classification grade 3 or higher) within 30 days after surgery.

RESULTS Of the 960 patients who were enrolled, 926 (96%) were included in the analysis. The mean age of participants was 63 years and 554 (60%) were men. Surgical site infection within 30 days after surgery occurred in 60 of 463 patients (13%) in the oral prophylaxis group and 100 of 463 (22%) in the placebo group (absolute difference -8.6%, 95% confidence interval -13.5% to -3.8%; relative risk 0.60, 95% confidence interval 0.45 to 0.80). The proportion of patients with deep infections was 4.8% in the oral prophylaxis group and 8.0% in the placebo group (absolute difference -3.2%, 95% confidence interval -6.4% to -0.1%). The proportion of patients with organ space infections was 5.0% in the oral prophylaxis group and 8.4% in the placebo group (absolute difference -3.4%, -6.7% to -0.2%). Major postoperative complications occurred in 9.1% patients in the oral prophylaxis group and 13.6% in the placebo group (absolute difference -4.5%, -8.6% to -0.5%).

CONCLUSION Among adults undergoing elective colorectal surgery, the addition of a single 1 g dose of ornidazole compared with placebo before surgery significantly reduced surgical site infections.

Statistical analysis

Assuming a 15% rate of surgical site infections with placebo,5 6 13 we estimated that enrolling 920 patients would provide 80% power to detect a 40% relative between group difference in the incidence of the primary outcome (ie, 15% in the placebo group and 9% in the oral ornidazole group),13 with a 5% two sided type I error. We inflated the sample size to 960 patients to account for a 5% loss to follow-up. As prespecified in the study protocol, one interim analysis was planned after the enrolment of the first 460 patients. The data and safety monitoring board did not recommend stopping the trial, and 960 patients were therefore included.

The planned approach to statistical analysis is published elsewhere.19 We analysed data in the modified intention-to-treat population, which was prespecified as all randomised patients who received a trial drug plus intravenous antimicrobial prophylaxis, with the exception of those who withdrew consent. We also analysed one per protocol population, which included patients from the modified intention-to- treat population except those with one or more major protocol violations.

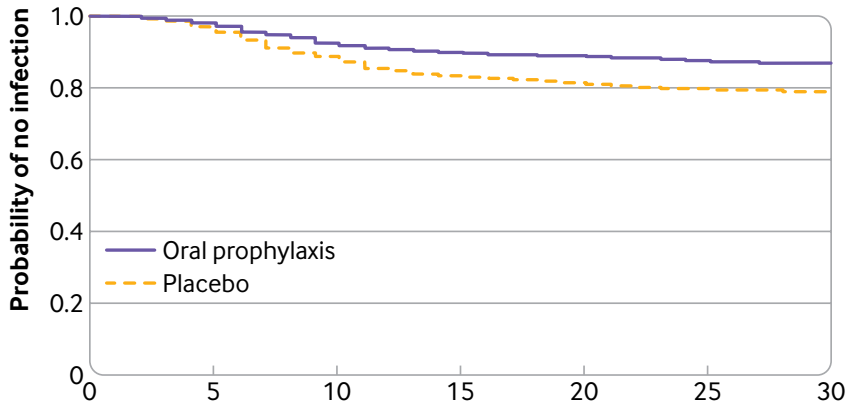
An unadjusted χ^2 test was used to compare the primary outcome between the two groups. Other binary outcomes were tested using an unadjusted χ^2 test or Fisher's exact test as appropriate. Results are reported as absolute differences and relative risks with 95% confidence intervals. Multivariable logistic mixed regression was used to identify prespecified covariates with a known association with the primary outcome (selected if the P value was < 0.10 in the bivariable analysis) in addition to the stratification variables. We assessed multicollinearity between variables by computing the variance inflation factor and using the Farrar-Glauber test. The Akaike information criterion and bayesian information criterion were calculated and used as model diagnostics to determine how well the model fit improved after the addition of covariates. Adjusted analyses were performed with the use of robust random effect Poisson generalised linear mixed model regression with robust variance for binary outcomes,27 multinomial logistic mixed model for categorical outcomes, and linear mixed regression for continuous outcomes, with study site as a random effect. Time to event was compared between the two groups using the Kaplan-Meier method. A marginal Cox proportional hazards model was used to estimate hazard ratios and corresponding 95% confidence intervals. The proportional hazard hypothesis was evaluated using the Schoenfeld test and plotting residuals.

We conducted two prespecified subgroup analyses of the primary outcome in subgroups with mechanical bowel preparation versus without and with colonic surgery versus rectal surgery. Interaction terms in the random effect regression model were used to test for heterogeneity of effect between subgroups.

A post hoc analysis was conducted to test for a difference in treatment effect during the conduct of the trial in relation to publication of the update to French guidelines (before versus after publication update). We also conducted a post hoc analysis to investigate a potential treatment effect resulting from non-compliance with bowel preparation. No correction for multiple testing was applied in the analyses of secondary outcomes or subgroups. Complete case analysis was performed for all outcomes. We did not compensate for dropouts. A two sided P value of ¡0.05 was considered to indicate statistical significance. All analyses were generated with the use of Stata software, version 15.0 (StataCorp).

Results

PRIMARY OUTCOME Surgical site infections within 30 days after surgery occurred in 60 of 463 patients (13.0%) with oral prophylaxis and in 100 of 463 patients (21.6%) with placebo (absolute difference -8.6%, 95% confidence interval -13.5% to -3.8%; relative risk 0.60, 95% confidence interval 0.45 to 0.80). Supplementary table S2 shows the results of associated univariable and multivariable analyses. The result was unaffected by adjustment for stratification variables and covariates (adjusted relative risk 0.62, 95% confidence interval 0.44 to 0.46) (see supplementary table S3). Similar results were obtained in the per protocol population (see supplementary tables S4-S6). Figure 2 shows the times to surgical site infection.



Oral prophylaxis	463	455	429	418	413	407	403
Placebo	463	450	412	387	378	370	363

Fig 2 | Kaplan-Meier probability of surgical site infection (modified intention-to-treat population). Raw data for the Kaplan-Meier probability of surgical site infection were censored at 30 days after surgery (hazard ratio with oral prophylaxis versus placebo 0.57, 95% confidence interval 0.43 to 0.78). The Cox proportional-hazards model was unadjusted

Q's (individual-level data can be reconstructed from Fig 2)

1. How many patients were lost to follow-up before day 30?
2. How many cases of infection had been averted by day 5? 10? 20? 30?
3. What was the point of the hazard ratio mentioned in the Figure legend? Does a constant-over-the-30-days HR make biological sense? Explain.