Later is not necessarily better: limitations of survival analysis in studies of long-term drug treatment of psychiatric conditions

Joanna Moncrieff [ID],[1] Janus Christian Jakobsen,[2,3] Max Bachmann[4]

[1]Division of Psychiatry, University College London, London, UK
[2]Copenhagen Trial Unit, Centre for Clinical Intervention Research, Copenhagen University Hospital, Copenhagen, Denmark
[3]Department of Regional Health Research, The Faculty of Health Sciences, University of Southern Denmark, Odense, Denmark
[4]University of East Anglia, Norwich, UK

Correspondence to:
Dr Joanna Moncrieff, Division of Psychiatry, University College London, London WC1E 6BT, UK; j.moncrieff@ucl.ac.uk

Abstract

Survival analysis is routinely used to assess differences between groups in relapse prevention and treatment discontinuation studies involving people with long-term psychiatric conditions. The actual outcome in survival analysis is 'time to event', yet, in the mental health field, there has been little consideration of whether a temporary delay to relapse is clinically relevant in a condition that can last for decades. Moreover, in psychiatric drug trials, a pattern of elevated early relapses following randomisation to placebo or no treatment is common. This may be the result of the withdrawal of previous treatment leading to physiological withdrawal effects, which may be mistaken for relapse, or genuine relapse precipitated by the process of withdrawal. Such withdrawal effects typically produce converging survival curves eventually. They inevitably lead to differences in time to relapse, even when there is little or no difference in the cumulative risk of relapse at final follow-up. Therefore, statistical tests based on survival analyses can be misleading because they obscure these withdrawal effects. We illustrate these difficulties in a trial of antipsychotic reduction versus maintenance, and a trial of prophylactic esketamine in people with treatment-resistant depression. Both illustrate withdrawal-related effects that underline the importance of long-term follow-up and question the use of tests based on time to event. Further discussion of the most relevant outcome and appropriate approach to analysis, and research on patient and carer preferences is important to inform the design of future trials and interpretation of existing ones.

Introduction

Survival analysis was initially developed to analyse risk of death over time, but is now used for the analysis of many categorical outcomes in health research, including relapse of mental health conditions. The outcome survival analysis assesses is the time to the outcome or 'event' in question. Common methods include Cox regression analysis, which produces a Hazard Ratio (HR) and is dependent on the assumption that the ratio between the hazard rates in the two groups is constant over time (proportional hazards assumption). The rank sum test does not make this assumption and tests the statistical significance of the difference in the overall survival time between the groups but does not yield an effect estimate. The Kaplan Meier method graphically describes survival over time.

Tests of differences between groups based on survival analysis are statistically efficient when the duration of participants' follow-up varies, and therefore commonly preferred to the alternative method of comparing the proportion of events at a particular time.[1 2] Limitations to do with censoring and low precision of the last stages of survival curves are well-recognised.[3] In the following paragraphs, we consider problems of analysing studies of drug treatment for relapse prevention in psychiatric conditions. These follow from the fact that survival analysis depends on the time it takes for an event to occur, combined with the potential for withdrawal of previous drug treatment to exert a differential influence on the risk of relapse in intervention trials that employ a discontinuation design.

Time to event is not necessarily the key outcome

Survival analysis is useful when the timing of the outcome is of importance and depends on people wanting to avoid early adverse events more than late ones. Tests based on survival analysis may show highly significant results if relapse is delayed by a few weeks, for example. However, the clinical relevance of a temporary delay in relapse in a long-term psychiatric condition that may last for decades has not been established, and statistically significant results are a questionable basis for implementing an intervention that may be of limited importance to patients. There has been no published research or discussion about whether patients, carers or clinicians consider a delay in relapse to be important, the minimum length of delay that would qualify, and how these views are influenced by the possible causes of relapse. Despite the widespread use of survival analysis there has been little consideration of its implications or how it compares with other approaches. In an analysis of relapse definitions in 81 trials of antipsychotics, for example, none of them considered whether time to relapse or overall risk was more relevant.[4]

Non-proportional hazards

If follow-up is short, differences between groups may be misleading, because short-term differences may not persist with longer term follow-up.
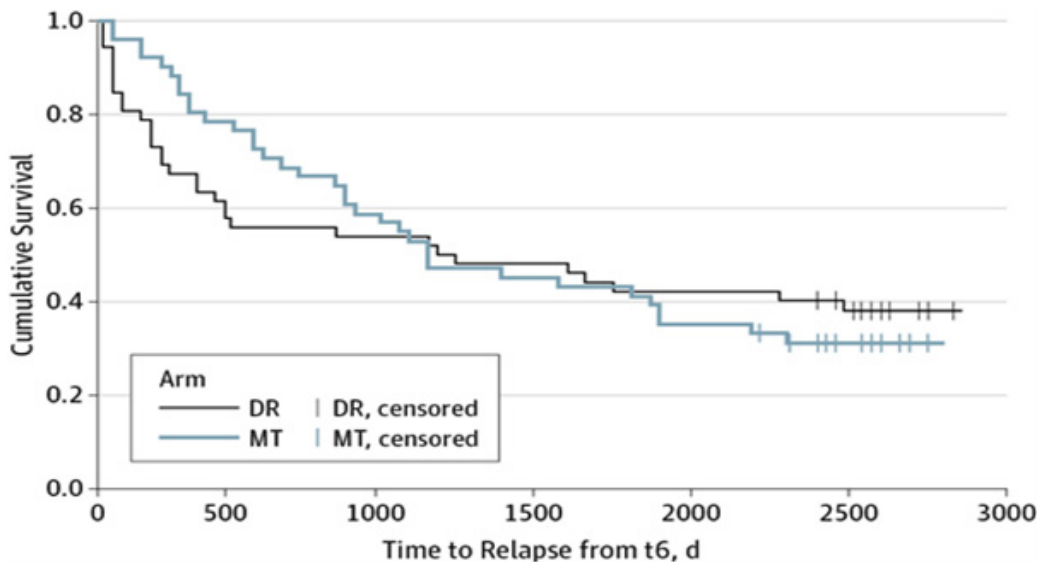
Figure 1 Time to relapse (days) among people with first-episode psychosis randomised to maintenance antipsychotic treatment (MT) or supported reduction (DR). Reproduced with permission from Wunderink et al.[8] Copyright (2013) American Medical Association. All rights reserved.[5] DR, discontinuation; MT, maintenance treatment.

In other words, it cannot be assumed that the HR or risk ratio between the groups remains constant over a longer duration. The difficulties caused by non-proportional hazards or risks when survival curves converge or cross are well-recognised by statisticians, yet frequently ignored when trials are reported.[5] In theory, a single HR is misleading in this situation, and Cox regression, which depends on the assumption of proportional hazards, should not be conducted. In practice, non-proportional hazards are common, but the proportional hazards assumption is frequently not tested and single HRs are often quoted.[5] Although the log rank test is technically correct in this situation, since it simply tests for a difference in the survival curves without assuming proportional hazards, it does not account for the pattern of the curves and its use forecloses discussion about the causes of varying hazards. Therefore, the use of the log rank test may also lead to misleading results, or at least results that are difficult to interpret clinically.[6]

Various statistical solutions have been proposed to manage situations in which non-proportional hazards are found.[6] [7] However, these are still based on 'time to event' analysis, and hence assume that delaying the event is the key desirable outcome.

Withdrawal effects in psychiatric trials

Trials of long-term treatment in psychiatry, including trials referred to as relapse prevention trials, employ a discontinuation design in which the withdrawal of previous treatment is compared with its continuation. Many trials of psychiatric drugs, including antipsychotics, antidepressants and lithium[8–11] reveal that people randomised to switch to placebo or no treatment show a high rate of early relapse that declines over subsequent months or years (eg, see figures 1 and 2). This high rate of early relapse is absent or less marked in those randomised to continue treatment, and hence there are substantial differences in relapse rates to begin
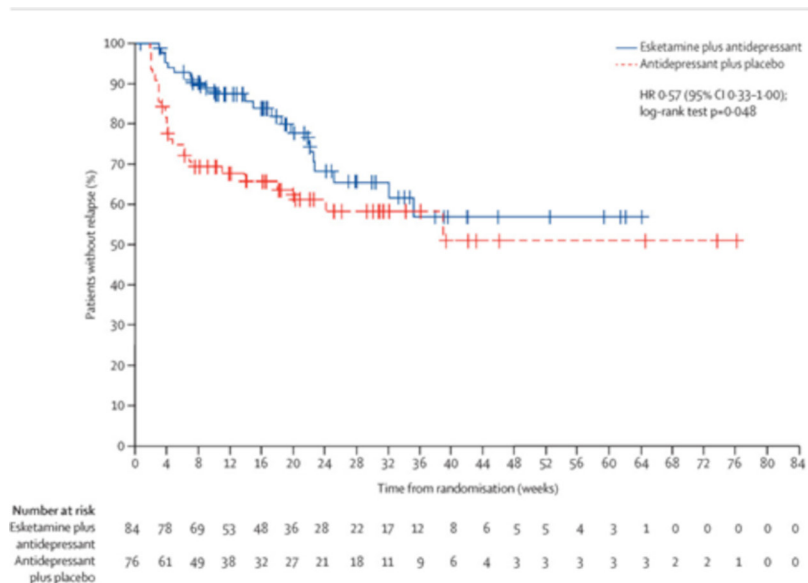

Figure 2 Time to relapse among people with treatment-resistant depression randomised to esketamine plus antidepressant or placebo plus antidepressant. Reprinted from Singh et al.[34] Copyright (2020) with permission from Elsevier.[21]

with, which diminish over time, as demonstrated in meta-analyses of antipsychotic and antidepressant relapse prevention trials, for example.[12–14] The reasons for the high rate of early relapse are debated and several possibilities exist.[15] It has been argued that it reflects a naturally high risk of relapse in some disorders, such as schizophrenia,[16] but the untreated risk in most psychiatric disorders, including schizophrenia, is not known. Where historical evidence exists, as in bipolar disorder, it suggests the underlying risk is lower than following discontinuation of drug treatment.[17 18] In a chronic condition, early relapses might also represent the re-emergence of underlying symptoms previously suppressed by the drugs, but this is only likely to explain a small number of early relapses, since most people do not have severe, chronic symptoms and long-term medication is generally prescribed for prophylaxis. Another possibility is that the withdrawal of previous treatment increases the risk of adverse outcomes above their 'baseline' level, due to the occurrence of withdrawal effects or of a relapse that is caused by the process of withdrawal.

It is well-established that psychiatric drugs of all kinds produce physiological withdrawal effects when they are stopped, especially if people have been using them for long periods.[19] These are manifested in physical and psychological symptoms, and may be mistaken for relapse, since symptoms overlap and there are no definitive ways of distinguishing the two situations.[20] Although withdrawal symptoms are traditionally thought to be short-lived, accumulating evidence suggests they can sometimes be protracted over many weeks or months.[21–23] In addition to this, withdrawal may itself precipitate relapse of some conditions, including schizophrenia or psychosis and bipolar disorder, elevating the risk of relapse for several months.[19 24 25] Studies of lithium treatment in people diagnosed with bipolar disorder, for example, show that the risk of developing an episode following the discontinuation of lithium is higher than it was prior to lithium being started.[10 26 27] The fact that some evidence suggests relapse is less likely with gradual compared with abrupt discontinuation also supports the possibility that it may be precipitated by medication withdrawal.[28]

Survival analysis in the presence of withdrawal effects

Several commentators have highlighted how withdrawal effects confound the interpretation of relapse prevention studies.[15 19 24 25] These studies may not, in fact, provide reliable data about the benefits of starting long-term medication, only about the adverse effects of stopping it. Nevertheless, trials of treatment discontinuation are valuable since many people are established on long-term treatment that may not be beneficial or that they want to stop. Whatever the purpose of a treatment withdrawal trial, the pattern of adverse outcomes following randomisation is important to understand. Survival curves can be helpful in illustrating such patterns if follow-up is long enough for withdrawal effects to evolve and dissipate. However, a single, global test of the difference between those who continue on treatment and those who withdraw over the whole period based on the time to relapse obscures these effects. This is important because the occurrence of withdrawal effects may affect how people view the desirability of remaining on, or coming off treatment. Moreover, such tests are often presented as if they are equivalent to tests of cumulative differences in risk at the end of follow-up. In a situation of constant hazards this is the case, but where hazards rates diverge or cross, results of tests based on survival analysis may conflict with tests of the eventual cumulative risk ratio. As figures 1 and 2 illustrate, if survival curves show an early divergence and then meet, tests based on the overall difference of survival curves will indicate a positive effect for the treatment that produces the

fewest early relapses, but tests based on the cumulative risks at later follow-up may show no difference. Two examples illustrate these arguments.

Example 1: antipsychotics and relapse prevention in first-episode psychosis

A trial conducted in the Netherlands comparing antipsychotic discontinuation with maintenance treatment in people with a first episode of psychosis followed people up initially at 18 months and then again after 7 years.

At the 18-month follow-up, the survival analysis revealed a constant HR, which was estimated by Cox regression analysis and indicated that the risk of relapse was increased by 2.3 times in the group randomised to discontinuation.[29]

Data from the 7-year follow-up, however, showed that survival curves converged at around 3 years and then crossed; therefore, the proportional hazard assumption of Cox regression would not be fulfilled (figure 2).[8] A log rank test would likely indicate an advantage for the maintenance group, however, because of the more frequent earlier occurrence of relapses in the discontinuation group.

As discussed, there is no evidence about whether people value a delay in relapse in this situation, and if they do, what duration of delay would balance out the considerable adverse effects associated with antipsychotics; furthermore there is evidence from the long-term follow-up of this particular trial that antipsychotics may impair social functioning.[8] Although antipsychotics were reduced more gradually in this than in other studies, the fact that the excess risk of relapse occurred early on in the group randomised to antipsychotic reduction suggests it may have been associated with the process of discontinuation. The study demonstrates the importance of long-term follow-up, and of not assuming that short-term outcomes are equivalent to long-term outcomes. The HR generated at the 18 month follow-up is not equivalent to the ratio of the cumulative proportions of people relapsing after 7 years, even though hazards were initially proportional. Use of the log rank test would obscure this withdrawal effect. Therefore, the cumulative proportion of people relapsing at different follow-up points, including the final follow-up, would be a better test of the overall outcome in this situation. This is what the authors presented in the follow-up report.[5]

Example 2: esketamine for relapse prevention in treatment-resistant depression

A further example is provided by the results of a relapse prevention trial of esketamine for people with treatment resistant depression.[30] Although esketamine is a relatively new preparation, withdrawal symptoms following recreational use of ketamine (a similar drug) are recognised, and include low mood and anxiety, which may be mistaken for relapse.[31] Psychological factors may also precipitate relapse following withdrawal, since significant unblinding is likely to occur after switching to placebo (due to loss of the psychoactive effects of esketamine), leading to anxiety about treatment withdrawal.

The original analysis of this trial was performed using the Kaplan Meier method and the log rank test showed a significant difference between esketamine and placebo (p=0.003). Survival curves indicated varying HRs but did not cross. The maximum divergence of risk between the groups occurred within the first 8 weeks following randomisation, suggesting a likely withdrawal effect. Although the authors of the esketamine trial asserted there was no evidence of a withdrawal syndrome, no details were provided, and it was not explained how relapse was distinguished

from possible withdrawal.[30] Subsequent data from this study and from others suggest that withdrawal symptoms occur commonly, are similar to those reported with ketamine,[32] and may be interpreted as a relapse of depression.

A letter published in response to this trial pointed out that results were strongly influenced by an 'outlier' site that reported a particularly large difference between the groups. The author of the letter analysed the data excluding this site by comparing the proportions of participants relapsing in both groups over the whole trial duration using the Fisher's exact test.[33] This found no statistically significant difference between the groups (p=0.13). The authors of the original trial objected that this was not the best way to analyse the data and applied survival analysis, as they had done in the original trial report.[34] This revealed survival curves that crossed at around 9 months of follow-up (figure 2). The log rank test was used to compare the groups and indicated a statistically significant difference (p=0.048), and Cox regression was apparently used to estimate a HR.

As the proportional hazards assumption is violated because the survival curves converge, it is incorrect to use Cox regression to calculate an overall HR, but the log rank test could also be misleading, because it obscures the convergence of the survival curves after what appears to be an early withdrawal effect. Treatment resistant depression is a long-term condition, and no research has yet clarified whether delaying symptom recurrence for a few months would outweigh adverse effects or represent value for money. It is also important to recognise the likelihood of a withdrawal effect, rather than to assume that the difference between the groups is the result of the treatment per se. Evidence of a withdrawal effect has a bearing on the cost-benefit analysis of starting treatment, and is particularly important in view of the fact that acute trials of esketamine have not demonstrated a clinically relevant effect.[35]

Patient and public involvement

There was no direct patient and public involvement in the preparation of this article, due to the lack of funding. However, there is a high degreee of concern among patients and the public about withdrawal effects from psychiatric drugs and how their impact has been misunderstood and underestimated.[36] The article was partly inspired by the first author's discussions with patients who are concerend about the interpretation of randomised drug trials.

Conclusion

Both examples illustrate how survival analysis may be misleading because it obscures a possible withdrawal effect, and delaying time to relapse has not been established as a worthwile outcome. Both studies also underline the importance of conducting long-term follow-up, since the outcomes of different interventions and treatment approaches can vary over time. Further research is needed on how patients and carers value a delay in relapse, and further debate is required about whether time to relapse should be preferred over the overall risk of relapse, especially in view of withdrawal-related adverse effects. Until then, we suggest that survival analysis should not be routinely employed in trials of interventions aimed at relapse prevention in long-term psychiatric conditions. Statistical methods for comparing proportions, such as the χ^2 test or logistic regression, should be used instead, complemented by Kaplan Meir survival curves to illustrate the timing of outcomes.

Twitter Joanna Moncrieff @joannamoncrieff

ORCID iD
Joanna Moncrieff http://orcid.org/0000-0003-1214-6974

References

1 Cuzick J. The efficiency of the proportions test and the Logrank test for censored survival data. *Biometrics* 1982;**38**:1033–9.

2 Gail MH. Applicability of sample size calculations based on a comparison of proportions for use with the logrank test. *Control Clin Trials* 1985;**6**:112–9.

3 Stel VS, Dekker FW, Tripepi G, et al. Survival analysis I: the Kaplan-Meier method. *Nephron Clin Pract* 2011;**119**:c83–8.

4 Moncrieff J, Crellin NE, Long MA. Definitions of relapse in trials comparing antipsychotic maintenance with discontinuation or reduction for schizophrenia spectrum disorders: a systematic review. *Schizophr Res* 2019.

5 Kristiansen IS. PRM39 survival curve convergences and crossing: a threat to validity of meta-analysis? *Value in Health* 2012;**15**:A652.

6 Li H, Han D, Hou Y, et al. Statistical inference methods for two crossing survival curves: a comparison of methods. *PLoS One* 2015;**10**:e0116774.

7 Royston P, Parmar MKB. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol* 2013;**13**:152.

8 Wunderink L, Nieboer RM, Wiersma D, et al. Recovery in remitted first-episode psychosis at 7 years of follow-up of an early dose reduction/discontinuation or maintenance treatment strategy: long-term follow-up of a 2-year randomized clinical trial. *JAMA Psychiatry* 2013;**70**:913–20.

9 Baldessarini RJ, Tondo L, Ghiani C, et al. Illness risk following rapid versus gradual discontinuation of antidepressants. *Am J Psychiatry* 2010;**167**:934–41.

10 Baldessarini RJ, Tondo L, Viguera AC. Discontinuing lithium maintenance treatment in bipolar disorders: risks and implications. *Bipolar Disord* 1999;**1**:17–24.

11 Tohen M, Sanger TM, McElroy SL, et al. Olanzapine versus placebo in the treatment of acute mania. Olanzapine HGEH Study Group. *Am J Psychiatry* 1999;**156**:702–9.

12 Leucht S, Tardy M, Komossa K, et al. Maintenance treatment with antipsychotic drugs for schizophrenia. *Cochrane Database Syst Rev* 2012:CD008016.

13 Gilbert PL, Harris MJ, McAdams LA, et al. Neuroleptic withdrawal in schizophrenic patients. A review of the literature. *ArchGenPsychiatry* 1995;**52**:173–88.

14 **Zhou D**, Lv Z, Shi L, *et al.* Effects of antidepressant medicines on preventing relapse of unipolar depression: a pooled analysis of parametric survival curves. *Psychol Med* 2020;**74**:1–9.

15 **Moncrieff J**. Why is it so difficult to stop psychiatric drug treatment? It may be nothing to do with the original problem. *Med Hypotheses* 2006;**67**:517–23.

16 **Davis JM**. Maintenance therapy and the natural course of schizophrenia. *J Clin Psychiatry* 1985;**46**:18–21.

17 **Winokur G**. The Iowa 500: heterogeneity and course in manic-depressive illness (bipolar). *Compr Psychiatry* 1975;**16**:125–31.

18 **Harris M**, Chandran S, Chakraborty N, *et al.* The impact of mood stabilizers on bipolar disorder: the 1890s and 1990s compared. *Hist Psychiatry* 2005;**16**:423–34.

19 **Baldessarini RJ**, Tondo L. Effects of treatment discontinuation in clinical psychopharmacology. *Psychother Psychosom* 2019;**88**:65–70.

20 **Cohen D**, Recalt A. Discontinuing psychotropic drugs from participants in randomized controlled trials: a systematic review. *Psychother Psychosom* 2019;**88**:96–104.

21 **Davies J**, Read J. A systematic review into the incidence, severity and duration of antidepressant withdrawal effects: are guidelines evidence-based? *Addict Behav* 2018.

22 **Hengartner MP**, Schulthess L, Sorensen A, *et al.* Protracted withdrawal syndrome after stopping antidepressants: a descriptive quantitative analysis of consumer narratives from a large Internet forum. *Therapeutic Advances in Psychopharmacology* 2020;**10**:204512532098057.

23 **Ashton H**. Protracted withdrawal syndromes from benzodiazepines. *J Subst Abuse Treat* 1991;**8**:19–28.

24 **Hengartner MP**. How effective are antidepressants for depression over the long term? A critical review of relapse prevention trials and the issue of withdrawal confounding. *Ther Adv Psychopharmacol* 2020;**10**:204512532092169.

25 **Récalt AM**, Cohen D. Withdrawal confounding in randomized controlled trials of antipsychotic, antidepressant, and stimulant drugs, 2000–2017. *Psychother Psychosom* 2019;**88**:105–13.

26 **Suppes T**, Baldessarini RJ, Faedda GL. Risk of recurrence following discontinuation of lithium treatment in bipolar disorder. *Arch Gen Psychiatry* 1991;**48**:1082–8.

27 **Cundall RL**, Brooks PW, Murray LG. A controlled evaluation of lithium prophylaxis in affective disorders. *Psychol Med* 1972;**2**:308–11.

28 **Tondo L**, Baldessarini RJ. Discontinuing psychotropic drug treatment. *BJPsych Open* 2020;**6**:e24.

29 **Wunderink L**, Nienhuis FJ, Sytema S, *et al.* Guided discontinuation versus maintenance treatment in remitted first-episode psychosis: relapse rates and functional outcome. *J Clin Psychiatry* 2007;**68**:654–61.

30 **Daly EJ**, Trivedi MH, Janik A, *et al.* Efficacy of Esketamine nasal spray plus oral antidepressant treatment for relapse prevention in patients with treatment-resistant depression. *JAMA Psychiatry* 2019;**76**:893–903.

31 **Morgan CJA**, Curran HV, **Independent Scientific Committee on Drugs**. Ketamine use: a review. *Addiction* 2012;**107**:27–38.

32 **Wajs E**, Aluisio L, Holder R, *et al.* Esketamine nasal spray plus oral antidepressant in patients with treatment-resistant depression: assessment of long-term safety in a phase 3, open-label study (SUSTAIN-2). *J Clin Psychiatry* 2020;**81**. doi:10.4088/JCP.19m12891. [Epub ahead of print: 28 04 2020].

33 **Turner EH**. Esketamine for treatment-resistant depression: seven concerns about efficacy and FDA approval. *The Lancet Psychiatry* 2019;**6**:977–9.

34 **Singh JB**, Daly EJ, Mathews M, *et al.* Approval of esketamine for treatment-resistant depression. *The Lancet Psychiatry* 2020;**7**:232–5.

35 **Horowitz MA**, Moncrieff J. Are we repeating mistakes of the past? A review of the evidence for esketamine. *Br J Psychiatry* 2020;**2**:1–4.

36 **Guy A**, Brown M, Lewis S, *et al.* The 'patient voice': patients who experience antidepressant withdrawal symptoms are often dismissed, or misdiagnosed with relapse, or a new medical condition. *Ther Adv Psychopharmacol* 2020;**10**:204512532096718.