

Learning from our Mistakes : How the “Wrong” Result can set you “Right” in Community Psychiatry Research*

Prof. Tom Burns

Professor of Social Psychiatry
University of Oxford
Department of Psychiatry
Warneford Hospital

It’s a real pleasure and honour to be here and I’ve read about Gerald Chao’s career and what he has done. I’m deeply honoured to be speaking in his name today. One of the things I learnt in my time in Hong Kong is that Hong Kong is a very international culture. You look to the East, you look to the West but what you don’t do is ignore the world around you. And that’s particularly relevant to the lecture that I am going to give you tonight.

What I’m going to talk about, I don’t want you to get fixated on the details. I’m going to tell you a bit of a story because it’s a real story and it’s a story in which I learned things I didn’t expect to learn. And I’m glad, although at one stage in this story I was very disappointed indeed, deeply disappointed, with what happened in our research study. In retrospect I think I learned more from things not turning out the way we wanted them to be than if they had turned out the way we wanted them to be. Hence the title “When the ‘wrong’ answer can give the ‘right’ results”.

I always like this quote – and many of you will recognize this distinguished man, this is Sir John Maynard Keynes an English economist who advised the English and the American governments on their economic policy in the late 1930s and then again at the

Bretton Woods Conference directly after World War II.

What is interesting he was giving this advice, a very influential man, and both governments listened to him, and somebody said to him in 1945, “But you are giving the opposite advice which you gave in 1936.” John Maynard Keynes was a rather pompous, rather aristocratic man, and he said, “When the facts change, I change my opinions. What, Sir, do you do?” And that to some extent is the keynote of this lecture. We may have strong views about things, we may have strong commitment to things, but if the circumstances change, if the facts are different, you have to be prepared to swallow hard and accept that maybe you were wrong. And in part of this story you will see very clearly that I was wrong and the research showed that I was wrong and it’s important when the research says that you are wrong, to listen to the research.

Mental Health Services Research

If we’d been talking about community mental health care 20 years ago, we wouldn’t have bothered about research at all. We’d have looked for experts and a list of people who described what they did and we would make a judgement about whether they did it well and

*Transcript of the presentation at the The 5th Dr. Gerald Choa Memorial Lecture of the Mental Health Association of Hong Kong delivered on 10th December 2010.

Prof. Tom Burns

whether their patients and their family doctors were pleased with it. Since the 1980s we've seen two important changes in medicine. One is the arrival of evidence-based medicine. We are expected as professionals to know the evidence and as far as we can to direct our practice according to it. Predominately that evidence is about individual treatment, but in the last 20 years we've had, for the very first time, a growing body of research into how to organise services. Before the 1980s we just developed them on a hunch, and if it worked well, we would replicate it. But since the 1980s, we've begun to develop a tradition of researching how we should deliver our care.

This mental health research, which I refer to as MHSR – Mental Health Services Research - not only has begun to grow in strength and status but it has clearly internationalised, and that in many ways is the mark of scientific health. You look abroad, you listen to other people. But in many ways it took a bit longer to get off the ground for health services research because how things are organised is often quite difficult to describe and to compare.

Basically we now have a tradition, a short tradition but a powerful tradition, of drawing on all the international research on how we organise our services. And so I would be a great advocate of that and I think it is the right thing to do, we should look around the world, we should not be parochial, we should not just say that I haven't done it so it doesn't work. That's not the way forward and that's clearly not the way forward in your culture.

But there are a couple of health warnings that have to go with international health services concerns. The first one which is often overlooked, particularly overlooked by politicians and sometimes overlooked by practitioners, is that in virtually every branch of medicine the findings, the effects size,

the power of an intervention, in a carefully controlled study is always much greater than in normal life. Whatever people find in their research study, you can't expect to do as well as them. They handpick their patients, they have got the best staff, all working for how to get their PhDs. Real life is not like that and in some other work I have just been doing on location grid limitation I've been looking at what is the difference of what researchers get and ordinary services get and I would think we've got to take at least a 50% drop in the effect size.

So the first thing to always remember is the difference between demonstration services and routine services. But the second one, and the one I really want to focus on tonight, is that particularly when you are looking at complex interventions in community psychiatry - context matters, culture matters. These are not disembodied drug molecules, these are interventions between people. Remember, all mental health care is an interaction between two people. Maybe more than two people but it always takes place between people. It is not independent of context and culture. Now clearly when we are doing research it is our job to make our research as reliable and context independent as we can. We do strive for that. In this sort of area I think we have to be quite careful about interpreting it. And to talk you through this I am going I am going to give you the story about assertive community treatment.

Psychiatric Assertive Community Treatment

Now assertive community treatment, originally called PACT, was a clinical practice described by Glen Stein and a social worker called Mary Ann Test in Wisconsin in the USA. A series of complications in 1982 had enormous impact and it was a natural experiment in which they had to close a ward and had staff with no jobs to do and they were fed up with discharging patients from that hospital who came back in within three to

How the “Wrong” Result can set you “Right” in Community Psychiatry Research

six months as ill as they were the first time. So they thought that they must be able to do better than this. What they did was set up a service using their staff to do intensive case management. Each full time worker only had ten patients, they followed them up daily, if need be, weekly at least. They did everything

that they could to help the patients stay well and particularly and radically they didn’t ask the patients to come to the clinic – they went to the patient. They had regular team meetings, they emphasised a lot of disciplinary approach and emphasised being available 24 hours a day.

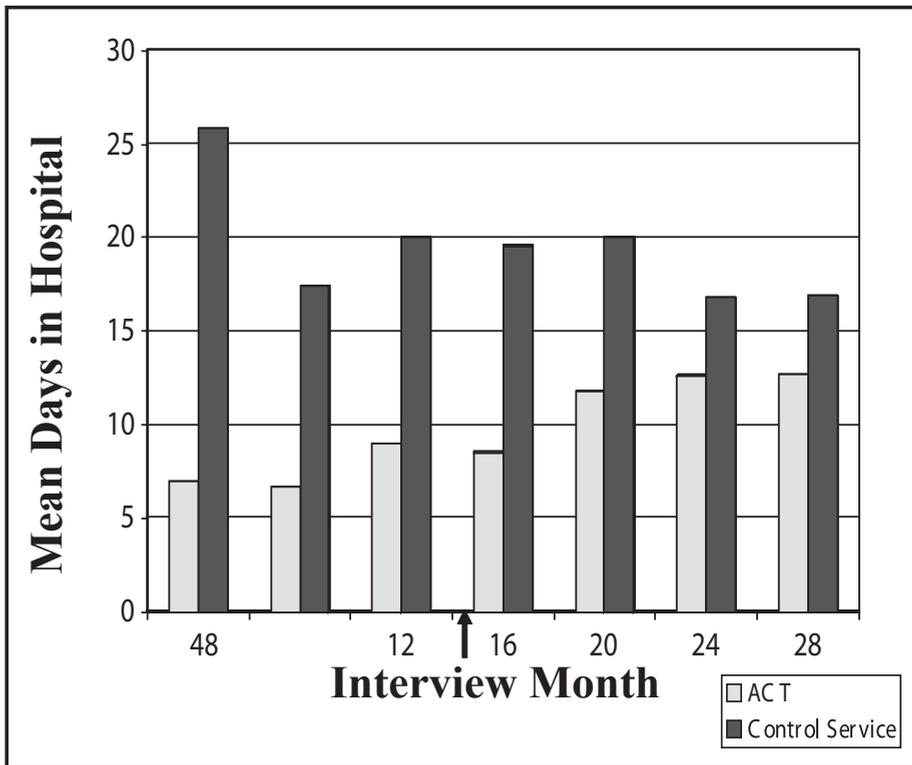


Figure 1 : Stein & Test 1980: Hospital use

This service was established in about 1979 but the important thing about it was that at the point at which they established it they did a random controlled trial of it. A very small trial, by current standards far too small - you wouldn’t probably get this published now. It was only 126 patients - 60 went into one arm, 60 went into the other. One arm got the normal support, which was nothing, and they came back to hospital when they needed it. Others were followed up in this way and they had fantastic results. They reduced hospitalisation massively. They could have improved social

functions, some of these patients got drugs. Depending on how you read it they either got the same symptom outcome or slightly reduced symptom outcome. They seemed to have got these patients to stay well out of hospital actually for less money, or certainly for no more money, than the sort of neglect that they were previously getting as costing because they saved so much on hospitalisation.

This is their finding on hospitalisation as they followed up their patients for up to 14 months and they had low level

Prof. Tom Burns

of hospitalisation, those who didn't get the intensive management - lots more hospitalisation.

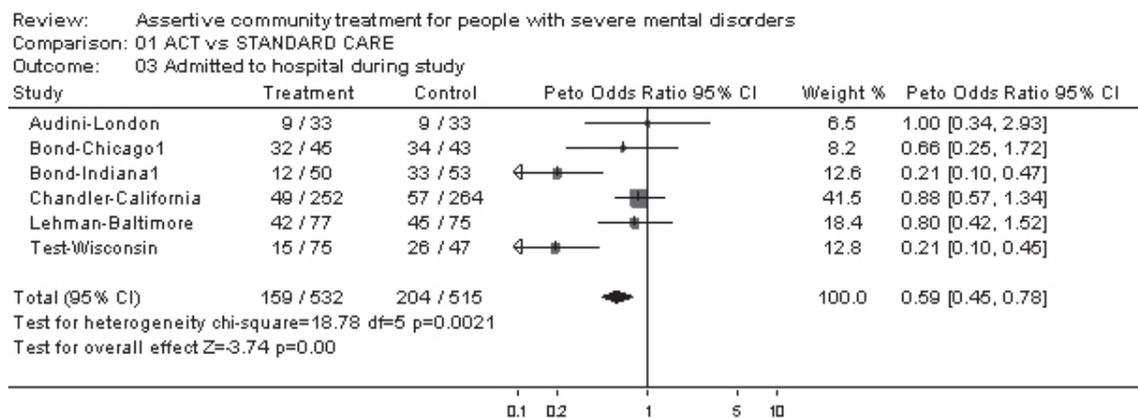
The interesting thing to note here (and this is an unexpected finding – because this was meant to be a 14 month study, and the money ran out, but they followed them up just out of interest) is that actually once the service stopped – the service stopped there – the advantages wore off, because actually Glen Stein thought of this as a training programme and if it wasn't a training programme the advantages would have stayed after here – but they didn't, so clearly it's a supportive service.

PACT was enormously successful and it became successful because of those wonderful results that reduced admissions by 60%. It actually seemed to save money and also a very important thing about it was that it was very well described in the papers they published - which was unusual at the time, sadly still unusual - they described what they did in good detail. So that if you looked at their research papers you would say - we'll do the same thing and test it here. So it had enormous impact both clinically and also in terms of the research

community because people thought they could replicate it, and they did.

So it was rolled out internationally and I think you will be familiar with it so I won't dwell on it but basically this a concrete review of ACT. What it shows here is that taking a bunch of studies together and adding them up, they'll all show this is the figure that matters, if it's not touching this line statistically significant improvement for one arm or the other. What this shows is that ACT significantly reduces hospitalisation and a similar examination of standard case management shows that it actually significantly increases hospitalisation. So if you put those two together we can say that ACT is an intervention that is bound to reduce a patient's need to be in hospital. The result of this meta-analysis - and I saw that my predecessor at this podium Louis Appleby and I don't know how much he talked about it - in 1999 we introduced a national service framework which drawing almost exclusively on this set of research said that we should disband our community mental health teams and work to assertive outreach teams and that's what was done.

Table 1
ACT vs Standard Care Hospital Admissions



How the “Wrong” Result can set you “Right” in Community Psychiatry Research

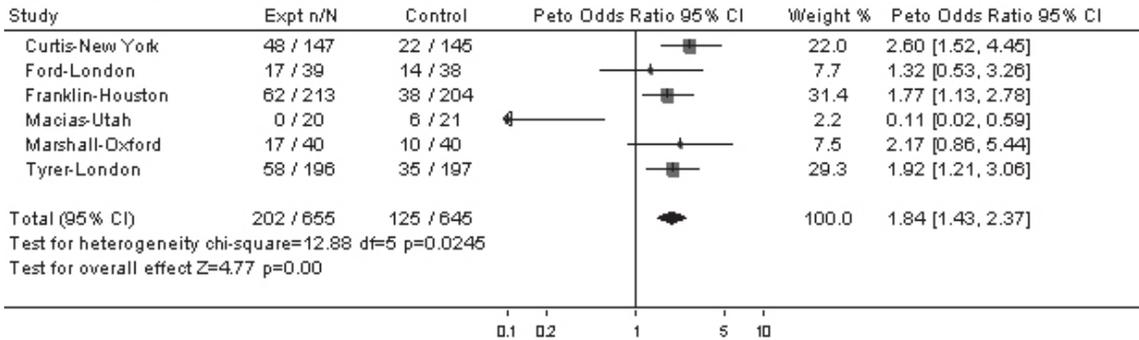
Table 2

Case Management vs Standard Care Hospital admissions

Review: Case management for people with severe mental disorders

Comparison: 01 CASE MANAGEMENT vs STANDARD CARE

Outcome: 03 Admitted to hospital during study



UK 700 Study

There was a problem with that and that’s where it starts to get tricky because at the time that this was all being published and the national service framework was being conducted, I and four other professors of psychiatry from Manchester, Maudsley, St. Mary’s and St. George’s in London, were conducting a really very rigorous and extensive randomised control trial of this approach. It was called the UK 700 Study - it was going to be called the UK 800 Study but we failed to recruit the 800 patients, but basically we recruited 700 patients across four sites and we followed them up for two years doing what was becoming and enforcing. This was a very staggering experience. At exactly the point that we were producing our results the government had insisted that we establish 300 assertive outreach teams. You can only establish 300 assertive outreach teams by taking staff from other mental health teams. And we were publishing a massive study in the Lancet, very rigorous, very carefully reviewed. We showed, and it was a shock to all of us, no difference. It didn’t just show that the difference wasn’t statistically significant or clinically significant; it showed no difference at all, not a day different between 350 patients in

intensive case management and 350 patients in ordinary CMHTs. Both were highly selected patients for being very severe and having lots of current health problems and both of them had 72 days average of hospitalisation over two years. It wasn’t 70 versus 80 or anything like that – it was 72 in both. So this was a real problem.

And at the same time Graham Thornicroft was publishing a study called ‘PRISM’ which had about 560 patients in it. It wasn’t scientifically rigorous but it had a fantastic advantage being epidemiologically based – testing exactly the same question and finding exactly the same answer – no difference. So it was difficult, it was embarrassing, because it was a shock. It was disappointing for me because I, as you heard if you listened to that very long introduction to me, I established the first assertive outreach team in Britain and I know, I had been running it. I was a fan, I was writing books on it, I thought it was the way forward. Bang, result - found no advantage. And I even helped the state act when that happened. I’d been peer reviewing for about a year before this, serious articles from America, Australia and Europe. It was much smaller and none of them could find the differences. Actually when you looked at them carefully

Prof. Tom Burns

we found that outside the U.S. and one study by John Bolt in Australia, nobody had found this difference, and we were convinced up until this point that it was absolutely certain.

When you get it wrong - have we got the same results as everybody else? I'm not sure that any of you would be able to do another trial. We would all have been convinced the best way to look after people with severe psychotic illness in the community is to have very tiny case loads – one to ten, to be available 24 hours a day, to have daily meetings, to follow the PACT model exactly. This is what is referred to as the high model fidelity and of course when we didn't find the result there was a wave of criticism of the study and most of that criticism was the reason we didn't find a difference. This is a legitimate criticism, in any complex intervention you always have to check – they said the reason you didn't find a difference is that you didn't do it right. Actually how you did the intervention, you didn't do it the same way these other people did it. You know that in psychotherapy research it is very important to test that people do deliver what they say they're delivering because this is all based on human beings and their motivation and their consistent practices. It's not a drug that can guarantee its consistency.

Now it just happened that we had someone doing his PhD. with me called Mathew, and what he was doing his PhD. on was the process of this case. We had kept incredibly detailed process notes of our service and we had friends in New Hampshire – a chap called Bob Drake who has done the New Hampshire assertive community treatment study - and he had recorded data and they had found a positive result and so we went and we looked at what they did and what we did and we looked at it together so there was no bias and we found that we had done exactly the same thing. So you couldn't say that we'd failed to find a difference because we just hadn't done the intervention right.

Can wrong results give the right answers?

So we were presented here with an intellectual dilemma. There's something like 20 studies of ACT out there. The bulk of them are saying that it should substantially reduce hospitalisation, but only two large British ones and one small British one by Frank Hollard. I mean no study is ever perfect – but the UK 700 one was extremely rigorous and the PRiSM one was extremely representative. None of them were profoundly flawed, none of them were knowing what the international evidence was telling us. So you can either just say, “well, bad luck, there we are, let's move on, let's do something else”. But you can put a lot of years into this and I was really quite perplexed by it. It became very important to me to try and understand why this wrong result – the fact that we hadn't reduced bed occupancy - happened. So basically I am going to tell you how we dealt with it. I think we learned three things in this process.

We learned why it didn't work, because I have to say the findings from UK 700 study didn't stack up but not from mistake. We actually confirmed that ACT doesn't reduce admissions in the UK, but I think we were able to find out why it doesn't – that's the first thing. But that's an interesting academic question. This is where it is good to get conflicting results. If all your results point to the same direction you can feel confident about what you are doing. But if your results are slightly varied a little bit or they contradict each other, although that may reduce your confidence it has one fantastic advantage, it gives you an opportunity to get below the surface. With slightly different settings it may give you a chance to understand. Bear in mind ACT is a very complex psycho-social intervention, and the Americans who developed it always insist that you must do it exactly as it's described on the tin; you mustn't vary anything.

The problem is if you never vary anything, you will never improve. So because we've

How the “Wrong” Result can set you “Right” in Community Psychiatry Research

got this variation, we can try to find out why it didn't work in terms of reducing hospitalisation, but couldn't we then use the variation to get a better understanding of what was important? Lastly I think that the whole exercise helped me and certainly a number of my colleagues sharpen up our scientific approach to community psychiatric research. Indeed, psychiatric research is not easy. Sometimes people say because it's so difficult let's take an easier route and not do difficult things like ACTs. I would argue we've got to do these ACTs, but this small journey taught us a little bit more about how to do those ACTs better. Those are the lessons I am here to share with you today.

There is a lot of argument going on about the failure of the UK 700 and the PRISM Study to reduce hospitalisation. Eventually it got down to a bit of a 'Punch and Judy show'. British psychiatrists - I don't know to what extent this is popular in Hong Kong - like setting up debates. We get academic A who believes this thing and academic B who says that, and we get them to argue in public and it's quite fun because you only give them eight minutes and they have to tell jokes and try to make the other one look a little silly, but it forces you to think about it.

We ended up with me and a chap called Max Marshall being continually asked to do this. Essentially we both had contradictory positions. Max Marshall said the reason British ACT teams do not reduce hospitalisation is because we do it badly - the model fidelity is poor, we don't actually follow the model well enough. Glen Stein used to say that it's an all or nothing model, you can't vary anything, the whole is greater than the sum of the parts. Rather a spiritualistic approach I would think. Scientists say you've got to try and look at parts.

My view is different. My view was a little perhaps rather chauvinistic because I've been to North America, I've been around the

services and there are two things that strike any European for those that look at North American mental health services. One, one is quite deeply impressed by the enormous quality of the academic institutions - of money, and of resources and of intentions to research and some of the demonstration sites which are absolutely perfect. But we are equally impressed by the complete - well I am staggered - by the fact that they don't mind the fact that if you are not in that service you don't get anything it; if are not in the service, well, that's up to you. They said this is our programme, and we get our programme right.

Now, I come from a culture, most of Europe, that says it's got to be for everyone. There's no point in having a fantastic service for 12 patients if there are thousands who are not getting anything. I always found that very striking in America. My interpretation of these findings was that it wasn't model fidelity; we've done the checking with Matthew's work, but that actually the control part was so poor on what was being offered.

Let me give you an example. I met Glen Stein. He's a most wonderful man, an aging Woody Allen, full of jokes - my daughter thinks he's Father Christmas. He's just wonderful - jokes, and card tricks, and everything like that. But, I remember him describing to me that at the point where they set up their PACT team, he said the standard discharge practice at that hospital was to give the patient, this is the chronic psychotic patient, as they were being discharged, a bus ticket to town and a letter in their hand to take to a doctor. It doesn't seem to me that there is much you've got to do to be better than that.

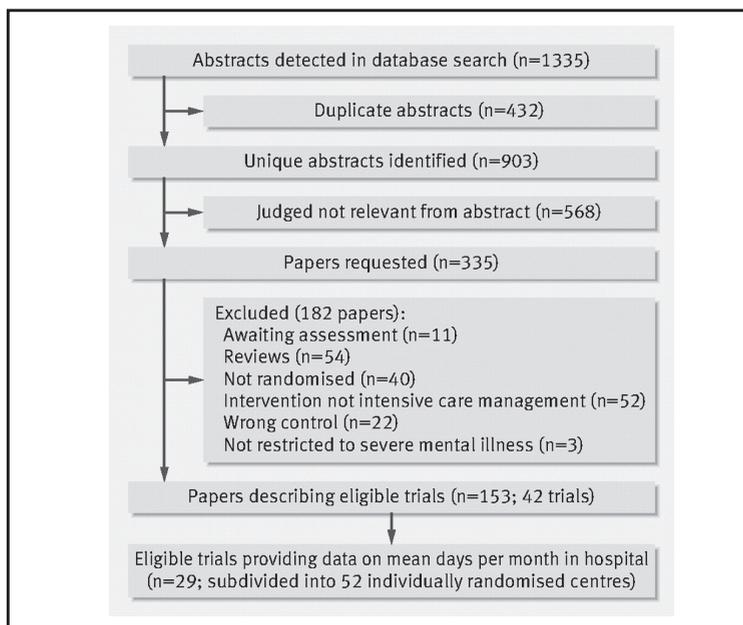
So I was convinced that the reason was it was the control of services. Max and I sat down and I said, "Look, it's quite fun doing these debates, but it's not getting us anywhere. Can we think of a way of actually testing this question? And this is what we did. This is an article published in the BMJ by Max and me and our teams. We came up with an idea of

Prof. Tom Burns

how we could explore why you were getting this massive reduction in hospitalisation in some studies and none at all in others. And what we did was to try and use intensive case management to reduce time in hospital systematically and meta-regression.

Meta-regression is the very technical statistical approach which I don't understand. It has some advantages over the statistical analyses seen. First of all you can include skewed data in it – you can't do that easily with a standard statistical analysis. Secondly, we went back to all the researchers and got

extra data. Sometimes there wasn't enough data in the papers to put them into proper review and we got the extra data. We could calculate standard deviations and the other thing that is special about a meta-regression analysis is that you can split multi-centre trials. Now normally multi-centre trials get their power by being all put together to give a big sample, but in fact clever statisticians say that if you've got a big study you can disaggregate it into sites, and in this meta-regression you get a more powerful test of what is causing the difference.



Burns, T. et al. BMJ 2007;335:336

Figure 2 : Metaregression of Intensive Case management studies
Process of identification of studies included in review

Max and I decided to bury the hatchet by doing this work together. So to some extent you've got the two proponents of the two conflicting explanations working together on an empirical answer to the question. There are lots of trials, and there are lots of patients - 8,000 patient data we've got in this study. This is a powerful test of what is causing the variation in the outcomes across the studies

and the number of studies we were able to break down with this extra power.

Although it's a very powerful test, 8000 patients, good science is not about going fishing. Good science is saying - what are our hypotheses at the beginning and testing them. When we spoke to the statisticians, they said that this data set would really only allow us to test four hypotheses with real certainty.

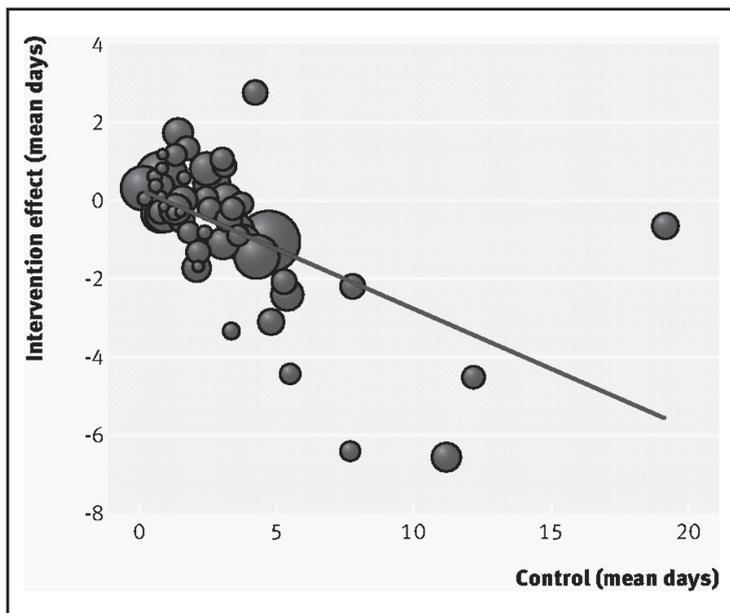
How the “Wrong” Result can set you “Right” in Community Psychiatry Research

The first two are technical ones, the ones for testing for publication bias and because frankly data study or size of study were major determinants in difference in outcome. Then you are talking probably about methodological problems, not about the data. But we put in here -this base line hospitalisation rates is my idea - the idea that actually the reason you don't reduce in some of the northern European studies is because we are already managing people well in the community to begin with and therefore it's hard to reduce, so that's what that stands for. And model fidelity was Max's idea, which was if you don't follow the description of PACT originally, the further you get from that, the less your outcome.

What did we find? Well we found an interesting answer and this is reassuring. We know it means that you can rely on the exercise, but both what I thought was important and what Max thought was important, both seemed to be important. So you could say it's a draw. I don't think it is a draw, actually I

think I won. But let me take you through why I won.

Impact of current references, i.e. what sort of normal community mental health services are you offering the people, to which you are comparing what matters at first. What you see here very clearly is if you've got the difference in days between the ACT patients and the non-ACT patients, the number of days in hospital per month over a 12 month period. The size of this group is the size of the population each of the aggregation of multi-site trials. Basically what you see here is – the number of days per month in hospital in your control services. So if you are managing your patients relatively well, say you are down here five days a month, that's the equivalent of what, 30 days a year, no, 60 days a year for very severe psychotic patients, you are only going to get a small reduction. You only get the big reductions because the further below zero you are, the bigger the reduction if you have got a service for which these patients tend to spend very long periods in care.



Burns, T. et al. BMJ 2007;335:336

Figure 3 : Metaregression of Intensive Case management studies
Control group mean v mean days per month in hospital.
Negative treatment effect indicates reduction relative to control

Prof. Tom Burns

Simon Bishop confirms this with considerable vigour that actually the main determinant, that's the slope of the line shows you how powerful the effect is, is why we weren't getting reductions in bed usage in ACT in the UK. This is because routine patients were only spending 20 or 30 days a year in hospital if we are doing what we are doing as well as ACT.

We also found that model fidelity also had an effect. There are different ways in measuring model fidelity for ACT. But basically you use a scale called IFACT which has 14 points on it because you can do it by

writing to researchers all from the notes; you don't actually have to watch the team do it. It has three scales on it, at least 14 items, which give you a score from zero to 14. Fourteen means that you are perfect, that you are doing services exactly science linked. Zero means that you are not doing anything right yet. Most teams are around 10, 12, 8 – they are close to it. To get those points you are going to have staff and you are going to have team membership, structure and organisation and these care practices you do actually have to be there to measure. But we could only measure these two.

Table 3

Scoring Criteria for the Index of Fidelity of Assertive Community Treatment

| Variable | Criteria for assigned score | | |
|---|--|---------------------|-------------|
| | <i>1/</i> | <i>ntermediate0</i> | |
| Client:staff ratio (CSR)C | SR < 10:1 | 2 - .1 x CSR | CSR > 20:1 |
| Team size (TS) | 7 < TS < 10. | 2 x TS - .4 | 10 < TS < 3 |
| Psychiatrist on team (POT) | POT > 13 hrP | OT/130 | |
| Nurse on team (NT) | NT > 30 hr/weekN | T/30 | 0 |
| Team primary therapist (PT)Y | es | -- | Otherwise |
| Separate site (SS) | Yes | -- | Otherwise |
| Shared caseloads (SC)S | core assigned based on clinical judgment | | |
| Daily team meetings (TM) | Yes | -- | Otherwise |
| Coordinator provides direct client service (DCS) | DCS > 20hrD | CS/200 | |
| 24-hr availability (AV)Y | es | AV brokeredO | therwise |
| Time unlimited services (TUL)Y | es | -- | Otherwise |
| In vivo contacts (IV)* | > 12.1/month | IV/12.10 | |
| Office contacts avoided (ie% office visits) (POV) | < 25% | 4/3 -P OV/75 | 100% |
| Total contacts (TC)* | > 18.5/month | TC/18.50 | |

Note: Unless otherwise indicated, scoring criteria are based on the mean value of experts' judgments for ideal model specifications from Study 1, assuming a caseload of 50 clients.

* Scoring criteria are based on model specifications from the Training in Community Living program

McGrew et al (1994) "Measuring the Fidelity of Implementation of a Mental Health Program Model", *Journal of Consulting and Clinical Psychology*

How the “Wrong” Result can set you “Right” in Community Psychiatry Research

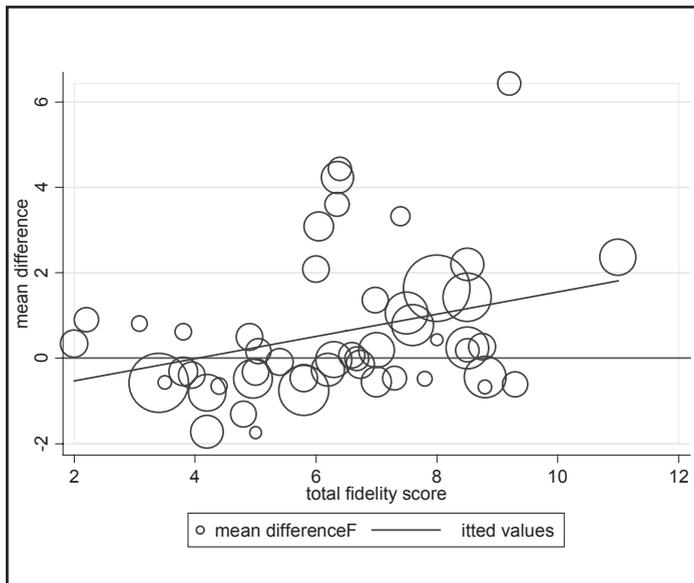


Figure 4 : Meta-regression of Fidelity v Reduction in IP days

So what did we find? Does model fidelity, does the closeness of fit of what you are doing measured by this scale, does it, is it associated with reducing bed occupancy? Basically the answer is – it does. If it didn't we would have a flat line. This is the same, this is the fidelity score getting stronger, and this is the mean difference. This is upside down compared

to the others, i.e. as you go up there you are reducing in patient care for the ACT group and you can see that the more closely you match the service, the more you're reducing occupancy, although the line of this curve isn't anywhere as near as strong as it is for controlled services.

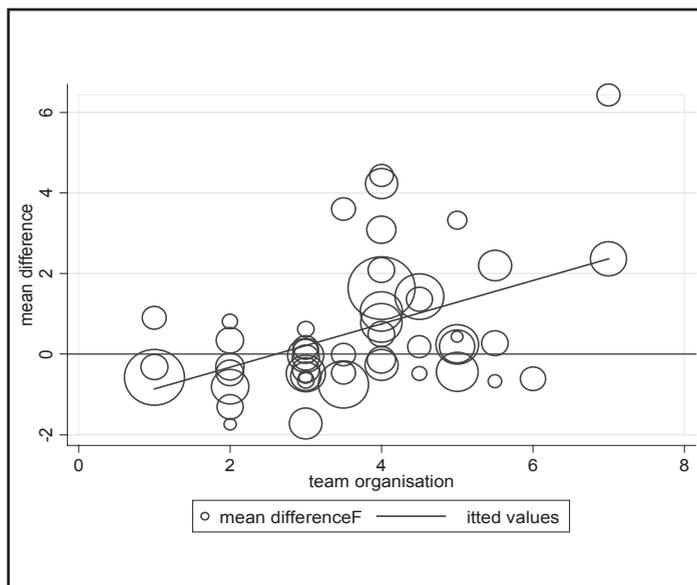


Figure 5 : M-R of Team organisation v Reduction in IP days

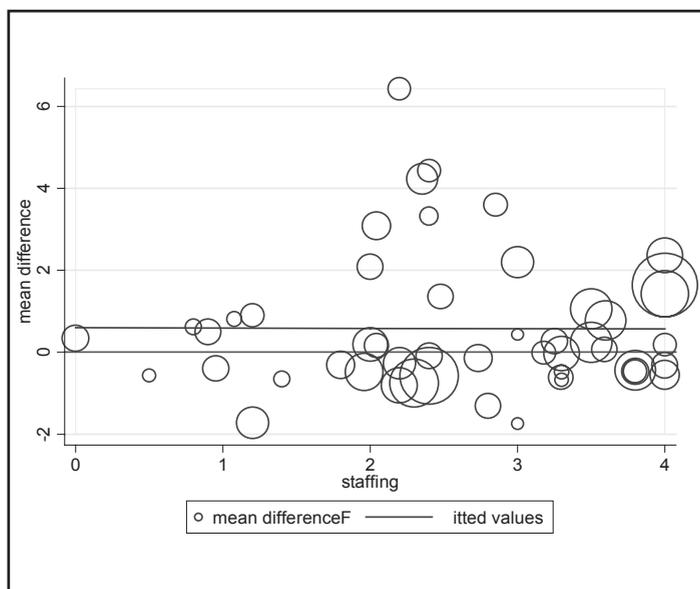


Figure 6 : M-R of Team staffing v Reduction in IP days

This is actually saying that we confirm that the very powerful effect is if your controlled services are very good, it's very hard to demonstrate an advantage for ACT. We are saying that actually if your ACT services are better than poor ACT services, you are getting, following this line, an increased reduction in hospitalisation. But the good thing about it, I found, is you can divide it into these two component scales. Now, this is team organisation which is - do you have multi-disciplinary working, do you need to get a team, do you have outreach? What we see here is quite a strong, slightly stronger curve, this is about two thirds of the scale and that is associated with a reduction. The most staggering one is this one.

The other component of the IFACT is the starting levels – one to ten case load, specialist workers, occasionally an invitation worker, diagnosis worker – have no effect at all. That means that the most expensive part of the service appears to contribute nothing. Organisation, style of thinking, concept of outreach, multi-disciplinarity, seem to be very important in reducing instability in our patients outside hospital, but the number of staff and the intensity of contact seems to have no effect

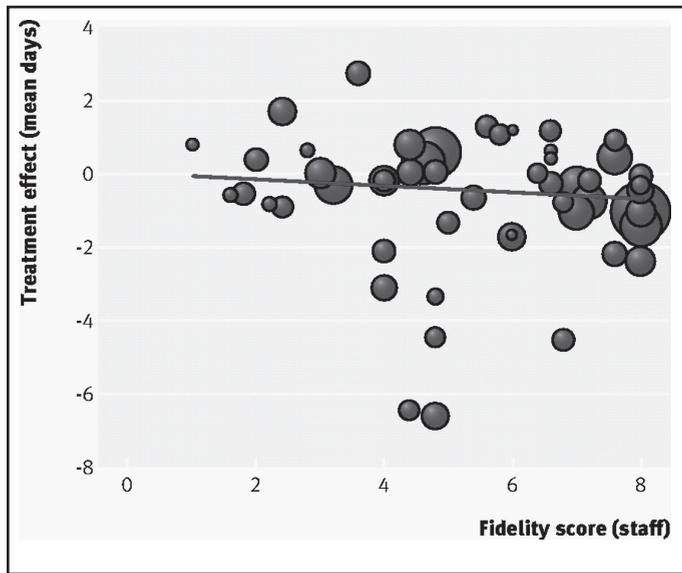
at all on these senses. Figure 7 & Figure 8

Community Psychiatric Service that works

I think that's a strange and surprising conclusion. I think we are forced to look at this evidence, to draw the conclusion that very small case loads of one to ten which are advocated in these sorts of outreach schemes are not justified in the case of the severely mentally ill. It's the most expensive part of the service and the disruptive part of the service and it appears not to contribute. Now we can have lots of discussions why that is, but the evidence seems fairly clear that going from 120 to one in ten, which is basically what we were comparing that, gives you no advantage – and that is a profound finding for anybody who has tried to fund mental health services where funding could be short. Essentially we could argue from that Assertive Community Treatment didn't reduce bed occupancy in the UK because it already has low bed usage.

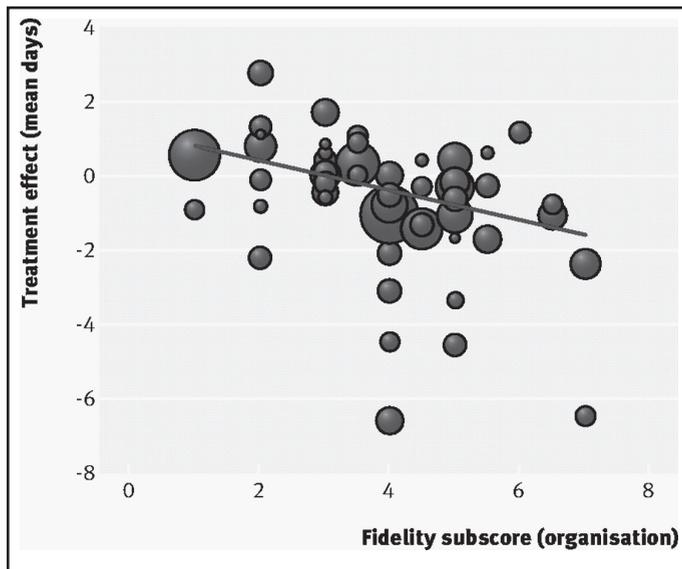
That's the first, if you want, of a shocking finding. We'd all gone into these studies, and every one of us in the UK 700 study was convinced we were going to find a massive reduction in hospitalisation. We didn't and we

How the “Wrong” Result can set you “Right” in Community Psychiatry Research



Burns, T. et al. BMJ 2007;335:336

Figure 7 : Metaregression of Intensive Case management studies
IFACT team membership subscore v mean days per month in hospital.
Negative treatment effect indicates reduction relative to control



Burns, T. et al. BMJ 2007;335:336

Figure 8 : Metaregression of Intensive Case management studies
IFACT organisation subscore v mean days per month in hospital.
Negative treatment effect indicates reduction relative to control

Prof. Tom Burns

were forced to rethink. It does appear that some of the core features of ACT don't actually add benefit and actually probably aren't necessary.

We did another study, which is nowhere near as scientifically rigorous as the one I have just shown you. In this one we wrote to all the people who have done home-based care for the severely mentally ill, who have done studies on it and done a randomised trial or a quasi-randomised trial and we asked them. We wanted to find out what they were doing. As I said at the beginning, people don't often describe studies well. We got a bunch of experts together and we said, "what do you think are the 20 most important things in good home based care?"

So we wrote to all these researchers and we said, "when you did that study that you published into home based care – usually showing reduced hospitalisation – can you just tick off which of these characteristics applied to your experimental team?" We're not interested here in your control team but in your experimental team, which of these things did you have. We got back over six hundred who did this questionnaire. Then we used the cluster analysis and a regression analysis to see if we could identify if there were any of these features of good home-based care which we thought were good. We didn't have any evidence of any of them clustering together or being very important.

Here is what we found. Eventually out of the 20 things only about six came up regularly. Wherever you look in the world where people were doing studies to support people's severe mental illness in the community, six things seemed to come out regularly as features of that approach. First of all they have smaller defined case loads - smaller here means one to 20, one to 25, not one to ten. But there are some services in the world, 20 years ago, services where people had 60 or 70 people. We said one to 20 is what we call a small case load. A manageable case load seems to feature quite often. Regularly visiting a home leads to looking as if they are the same but they are not quite. What this means is - do you have a

standard policy that you go and see patients at home rather than wait for them to come and see you? This was a question of what proportion of your contacts were in the patient's home because even if you have that policy you may only see 20 patients, 20 contacts a day or you may have 70%.

Multi-disciplinarity of the team is just to have three separate professional identities in it. We didn't say which they had to be. Most of them had nursing and medicine, some had psychology, some had social work (actually some American teams didn't have medicine) they would have a social worker and psychology and occasional nursing. But basically this is three types of discipline – professional discipline.

Now, this is an interesting question. We asked two things about the psychiatry. We asked how much psychiatric time is that, because actually that varies enormously. That didn't seem to come back as a regular feature, but what did come back as an important feature is how the psychiatrist worked in the team, because there are two ways of doing it which you see if you go round the world. One team functions without a psychiatrist but uses the psychiatrist as a resource if they need them. The form we use is much more. I've seen it in Hong Kong (it's certainly an English approach) which is that the psychiatrist is routinely a member of the team, part of the decision-making process all of the time, so it is integrated into the team. That seemed to be regularly reported.

And lastly, this one will be defined quite explicitly, we said, "Can your team deal with health and social coefficients in the team simultaneously?" Often that was associated with having a social worker in the team but it meant that you could deal with housing, you could deal with money, not just medicine.

And those are the six things that did seem to come up and I think we can have confidence to know that people who are going to be running a community-based service, the people sitting there doing this, are probably

How the “Wrong” Result can set you “Right” in Community Psychiatry Research

going to need most of these. In the regression analysis in which we tested it against the rate of reduction, these are the two that came out as most significant. Is it because you are able to stabilise people’s housing, accommodation, or is it because the social worker brings a more holistic approach to simply understanding people? But be that as it may these are the two that seem most important.

The second conclusion coming out of this surprising wrong result is that it forces us to understand why we’ve got a wrong result and it forces us to dig a bit deeper. We’ve got, I think, a slightly better understanding of what helps patients in the community, not just say ACT team helping you. We can now say that irrespective of the team, being visited at home, having your health and social care needs dealing with simultaneously and given due attention. So you don’t have to wait in a queue when you are referred to social services, not a disciplinary the psychiatrist being part of the team and manageable case loads, whatever that means. So we learnt that.

I’m now going to finish by just very briefly saying, “Did we learn anything about research methodology, did we learn anything from this surprising setback that can prove not just how we develop care, because those first two conclusions ought to help us target and deliver care more efficiently? Did we learn anything about conducting our research?”

Concluding remarks

Well, I suggest – I didn’t realise that this is such an English expression that most Europeans have never heard of the dodo. Have you heard, do you use this expression “dead as a dodo”? No, I should explain that the dodo was a bird in Mauritius that everyone thought was such a wonderful bird that they kept killing them and in 1864 they killed the last one. It exists – it’s stuffed in a museum in Oxford and it’s the last dodo. When you are talking about making something extinct – getting rid of something forever – we say you want it to be “as dead as a dodo”. So I am sorry about putting that in here.

But I began as a result of this. It highlighted to me a significant piece of sloppy thinking that we in the community of psychiatry research are particularly prone to and perhaps more than anybody else. That is that we refer to treatment as usual as our control comparative work. Now my Dodo Bird Society is dedicated to stopping having treatment as usual studies because - not that we shouldn’t have treatment as usual as a comparative – but to just call it treatment as usual tells you nothing. If you remember the first bit of that was pointing out that you had all these misunderstandings that arose because it was the treatment as usual not the intervention, the treatment as usual theory and we were just treating it as if it was the same.

Stefan Priebe and I had an article in ‘Psychiatrica’ 15 years ago, in which we wrote a proposal that journals should stop publishing articles in which you don’t at least devote as much time to describing your comparative breakdown as you do for your experimental breakdown. Now, I don’t think anybody read that article and certainly nobody changed their practice. But I recently had an editorial in the British Journal Of Psychiatry called ‘End of the Road for Treatment As Usual Studies’ – I don’t think it is the end of the road but you’ve got to keep making these points over and over again.

What was interesting was a recent publication from the REACT study. Now the REACT study was a study of assertive community treatment exactly like the study we were describing earlier – this is another one being done and comparing ACT versus CMHTs – published after three years, no difference. So what they wrote in the conclusion was – well maybe we didn’t go on long enough to find the difference. They did a long-term follow up and in the long-term follow up they found no difference. So their conclusions were, sadly - we have not been able to prove that ACT was superior to CMHTs.

If you’ve listened to what I have been saying up to now and you stop and think for a second, I was asked to review this article to

Prof. Tom Burns

see if it should be published. I thought that's interesting because actually they have not failed to show that ACT is better than CMHTs; they've actually shown that CMHTs are better than ACT because they've got exactly the same result for 60% of the resource. In any other context, if you've got the same clinical reward for 60% of the input you would say that's the right way to do it. What was interesting is (when I look back at our UK 700 Study) that was the conclusion we should have drawn but we didn't.

So to some extent I think one of the other things that we learnt from this exercise is that to do a random control trial you have to equipose when you go into it. You have to believe that you really do not know whether arm A or arm B is better. If you are convinced and certain, then it is unethical to do it. And what is interesting is that most of us can get ourselves into that equipose – we are also forgetting about calling it treatment as usual. But then when we come to analyse the results, we analyse the results not as if it is equipose, comparing treatment A to treatment B, we analyse it with a clear given conception that we are using the comparator group simply to compare, when in fact we are comparing two equally potentially effective treatments.

So I think there was a third finding there from this exercise, so having started off with a very disappointing result, we thought we'd failed with the UK 700 Study, we learned really why, I think, we learned quite convincingly why the UK 700 Study did not deduce it. We then went on further from that and we found out what we think were the effective ingredients.

And lastly I think we've underlined the need for more, not less rigour in community psychiatry research because it's fuzzy, because it's variable, because it's culturally contextual. That's not a reason for taking a softer experimental model, but I think it's a reason for having a much more rigorous experimental model and that there is no such thing as controlled service. It's a comparator service and it may have its strengths and

weaknesses, and if we don't know what they are, we don't know what conclusions we can draw from the study.

And so I think to finish I would suggest that what that means, if it means for us here gathered tonight in Hong Kong taking forward these initiatives and taking forward the question about being extremely alert and taking on board international research, is that we have to test it locally. We have to not assume because it works in Seattle, or it works in Camberwell, it will necessarily work in Hong Kong. And if we do research we have to take it seriously even if we don't like the results. If the facts change, we change our opinion.

摘要

從錯誤中學習：錯誤的結果如何令社區精神醫學研究的方向變得正確

在1980年代起，社區精神醫學的研究對社區精神科服務推行起著指導性的作用。講者在文章中分享了有關推行精神科「緊密社區治療」（簡稱ACT）研究的經驗。有見於美國有關ACT的研究顯示有助大幅減低入院比例達60%，講者於是在英國推行了一個大型的相關研究，參與的病人達700人。研究結果卻顯示ACT未能如期地降低病人入院的比率。講者分享他如何從研究上看似錯誤的結果，啟發他在社區精神醫學研究領域上走向正確的路。

Suggested Readings

Burns T, Greed F, Fahy T, et al (1999)
Intensive Versus Standard Case Management for Severe Psychotic Illness; A Randomised Trial UK 700 Group. *Lancet* 353, 2185 – 2189

PRiSM Psychosis Study (1998)
Br J. Psych 173, 363 – 427

Thornicroft G & Tanseller M (2004)
Components of a Modern Mental Health Service : A Pragmatic Balance of Community and Hospital Care. Overview of Systematic Evidence
Br J. Psych 185, 283 - 290