AUDIO INTERVIEW TRANSCRIPT

Sweetnam, Peter: transcript of an audio interview (31-May-2000)

Interviewer: Andy Ness
Transcriber: Jaqui Carter
Editors: Tilli Tansey, Hugh Thomas
Date of publication: 13-Feb-2017
Date of interview: 31-May-2000
Publisher: Queen Mary University of London
Collection: History of Modern Biomedicine Interviews (Digital Collection)
Reference: e2017052
Number of pages: 16
DOI: 10.17636/01019371

Acknowledgments: The History of Modern Biomedicine Research Group is funded by the Wellcome Trust, which is a registered charity (no. 210183). The interview was funded by a Wellcome Trust grant (059533; 1999-2001; awarded to Professor George Davey-Smith, Dr A R Ness and Dr E M Tansey), and its publication by a Wellcome Trust Strategic Award entitled “Makers of modern biomedicine: testimonies and legacy” (2012-2017; awarded to Professor Tilli Tansey).


Note: Audio interviews are conducted following standard oral history methodology, and received ethical approval from UCL (2000) and QMUL (reference QMREC 0642). Related material has been deposited in the Wellcome Library.

© The Trustee of the Wellcome Trust, London, 2017
Sweetnam, Peter: transcript of an audio interview (31-May-2000)*

Biography: Mr Peter Sweetnam (b. 1941) was a statistician at the Epidemiological Research Unit (South Wales) from 1966 until the final cessation of MRC funding in 1999.

AN: Andy Ness

PS: Peter Sweetnam

-------

A: Can I just start by getting you to tell me where and when you were born and a little bit about your early family life.

PS: I was born in a little place called Barn Green, just south of Birmingham, in 1941. My father was killed in the war two years later in 1943 and after the war my mother moved back to Cardiff, where she had been born and where she had two sisters and her own mother were still living. So I went to primary/junior school in Cardiff and then apparently, because it had always been my father’s ambition that his children were sent away to school, my mother managed to organize that and so I went away to what we called a public school in the middle of the wilds of North Devon, a college called Shebbear College, of which I have a mixture of mainly fond memories, but most of the time I rather wished I had been somewhere else. In the early 1960s I went to Bristol University where I did a degree in mathematics, and I then worked for what was then Bristol Siddeley Aero-Engines in Patchway [Filton] in Bristol for a couple of years as a technical engineer. This was at the time, I guess the first Labour government since the immediate post-war one, and they had cancelled a couple of Bristol Siddeley’s rather large engine contracts, a supersonic version of the Hawker Harrier engine and it looked as though they were going to cancel the Concorde at any second. It looked as though there were going to be large-scale redundancies on the basis of last in first out, and I thought I was going to get the chop pretty fast. So I looked around for something else to do, and I decided that I would do a one-year MSc course in statistics, back in Cardiff, and so after two years at Bristol Siddeley, I left and went back to Cardiff. In the event, in fact they didn’t cancel Concorde, and they made almost nobody redundant from Bristol Siddeley, but the die was cast by then.

So I went back to Cardiff where the course was a nine month’s tuition and examination course, followed by a three-month’s dissertation. And the course was taught by two people, one of whom had worked for a long time in industry with what was then British Nylon Spinners at Pontypool and had done a lot of theoretical work on quality control – he taught statistics as the mathematical theory of monotonic non-decreasing functions of the value between 0 and 1. The other half of the course was taught by Charles Rossiter, a medical statistician from the Pneumoconiosis Unit, and you would be quite hard pressed most of the time to realize that they were talking about the same subject. It was totally different, but Charles Rossiter got me interested. The pure mathematics, the mathematical theory, was really hard going I found. Charles Rossiter got me very interested and so I simply wrote to the Medical Research Council and said I was doing this course, did they have any jobs and if so could they suggest where I wrote. And I got a letter back from Archie Cochrane, who said ‘come and have an interview’. And so I wandered along to Richmond Road and had what was rather euphemistically called an interview, which mainly consisted of listening to Archie enthuse about what he was doing and I didn’t seem to do much other than nod and say yes that sounds

* Interview conducted by Dr Andy Ness, for the History of Twentieth Century Medicine Research Group, UCL, 31 May 2000. Transcribed by Mrs Jaqui Carter, and edited by Professor Tilli Tansey and Dr Hugh Thomas.
very interesting’ half a dozen times and I found that I was offered a job. And so I thought right, yes, here we go, and then basically there I stayed. I thoroughly enjoyed it by and large and in fact I started work immediately after the exam, before I had done the dissertation. I thought well, I can do this while I am working, but it took me another five years. I did end up doing it, but only after the university said the time has run out for this, either you get it in by such and such a time, or else we forget about the dissertation. So I started working in the Unit something like early June 1966 at which time there was one other statistician there, it was Hubert Campbell, who worked, his theoretical position was that he worked half time for the MRC and half time for the School of Medicine. I think the School of Medicine got the better end of that deal to put it mildly and so I was by and large the one statistician. But even then, particularly because of Hubert’s attachment to the medical school, but also because of Archie’s attachment to the medical school, a lot of the work that I did in the early days was rather with people in the medical school, a variety of people, than with the people in the Unit. I got involved in three or four clinical trials of various drugs in gastric ulcer and duodenal ulcer, with people like John Rose. I think the first paper I am an author on is one on fibroids and hypertension with one of the gynaecologists at the Heath Hospital. So that’s how I arrived and, as I say, basically thereafter that I stayed.

AN: And you worked, as you say, initially both doing MRC projects and things outside. In the first years what would be your main role, what would you be doing? What was the job of a statistician in the 1960s?

PS: It was mainly analysis of data. The work was put in front of you until one got a lot more experienced, when one got invited along to join the discussions about the design of the study in the first place. The analysis in those days was hard work, because we only had electromechanical calculators, punched-card sorters (Hollerith machines), so that even to do a multiple linear regression with two or three variables, if you had 200 individuals, was extremely time consuming. If you looked at the statisticians’ office then you would have found mountains and mountains of paper with numbers on and sums of squares on the ends of all the columns and rows. And then one checked them, double-checked them, cross-checked them, trying to make sure that you had got it right, because, of course, it was terribly easy to get them wrong. There were also purely practical things, such as the electromechanical calculators, which would jam regularly as clockwork if you tried to be a bit too quick with them.

AN: These electromechanical calculators, I don't think I have ever seen one. Could you describe what they were like.

PS: An electromechanical calculator would be about 18 inches square, with a moving carriage on top. The keyboard had long rows, of 0s, 1s, 2s, etc. You could enter a two-, or perhaps three- digit number at the left-hand end of the keyboard and another at the right-hand end. An instruction to multiply that number by itself would then produce, on the moving carriage, the square of the left-hand number, the square of the right-hand number and twice the cross-product. These squares and cross-products could be accumulated. This could work well, provided the original numbers weren’t too large and provided there weren’t too many of them, when the sums could overlap. But it was fairly easy to get this wrong. The only electrical assistance was to drive the carriage up and down, the rest of it was basically a mechanical operation. Very, very time consuming. I can't date this. Later we had what was probably the very first purely electrical calculator with a programming capacity, called a Monroe Epic. It consisted of a box that was rather like the electromechanical calculator, and another larger box that contained four large printed circuits. There were two versions of the Epic, one with 14 programming steps and another with 42 steps. We, I think, had the former. The machine had four registers in which numbers could be stored, but all arithmetical operations were done between registers 1 and 2. Registers 3 and 4 were simply non-accumulating stores. The calculations that could be done even with the 42 programming step machine were thus very limited as many programming steps were used to switch numbers between registers. These Monroe Epic calculators would have appeared around 1967/68. Computers were beginning to appear at about this time. I don't remember whether the Unit sent me, or whether I went of my own volition to what was then Newport Tech and did a FORTRAN programming course on an Elliot 803. That must have been in the late 1960s. At that time the only computer in the university as far as I know that was available for people like me to use was a Stantec
Zebra, which was a valve machine and where all the data and the programmes were inputted onto paper tape, a real nightmare of a problem to handle. It was a very time-consuming business and you were really very restricted in the analysis that you could do by the sheer practical limitations. It could take days to do a multiple regression with this equipment, unless there were very few subjects, and just one or two variables.

**AN:** Given that this was one of the first epidemiology units I suppose, and beginning to collect large numbers, did you feel that you were facing unique challenges or were there people with this expertise that you drew on?

**PS:** I don't remember. I don't know really. Well obviously there was Hubert Campbell around. He had obviously got a lot of experience of analysing data, but I think you really just restricted yourself to what was physically, practically possible. As I say they were really very, very simple analyses, and I think if you look at the early unit papers certainly up until the middle seventies, you won't find much other than a multiple linear regression with a couple of variables, and that's about the sum total of it.

You see if you think – I don't know whether I am jumping around now or not – Peter Elwood's first aspirin trial, which was published I think in the *British Medical Journal* in 1974. In the basement here [in Canynge Hall, Whiteladies Road, Bristol] you will find my hand calculations for the life tables and, literally, they were done by hand. There are sheets upon sheets of data, because we still didn't have access even then to a computer on which you could easily have done a life table. It's so difficult to try to get things chronologically correct, because I am sure that by then the university had an ICL System 4, and we used it for a few years by going over to Park Place in Cardiff, sitting down at the punch-card machines and punching out programmes and control cards etc. We first started doing that around about the same time as I was doing the analysis for the aspirin trial. Writing a programme for a life table analysis then was not the easiest thing in the world to do.

**AN:** So the data would come to you then as a pack of cards.

**PS:** Data arrived usually on bits of paper, sometimes it was decks of punch cards. We had some hand punches in the Unit, the nearest equivalent was the old individual credit card machines, with a block of metal and a little keyboard, and you would just slide the card in and type away and the card came out punched with the appropriate holes. There was also another machine, a verifier, whereby you’d put the same card in and you would type the same stuff in again, and it would stop if you hit a key that was different from the hole that was punched in the card. Staff like Janie Hughes used to do that. Those machines must have been available when I started, because I am pretty sure we would have used those on a Hollerith counter-sorter machine. The Hollerith could be set to look at a particular column and it would then run all the cards through and sort them physically into piles according to what was punched in that column. It was far from fool-proof, although a slick operator like Fred Moore, the person who mainly used the counter-sorter to sort the cards in piles, would pack them all altogether and lift them up and check that he could see through the hole to make sure that the machine hadn’t sorted any into the wrong place. The early machines just sorted the card, they didn’t count. Having counters on a machine was a later addition. I think the one we had when I first arrived would have had counters on it as well.

**AN:** So you would get some paper data, and then you would decide which variables to punch or do things, and you or someone else would punch it onto cards.

**PS:** I didn’t do much of the actual data punching. I don’t think that we ever had a large, half-electronic card punch machine like they used to have over in the university, banks and banks of them. I don’t know whether Janie Hughes used to go over to the university and punch cards over there or up to the Heath Hospital. At a later stage we used the university department of what was then called Medical Statistics and Computing, of which Hubert Campbell was initially senior lecturer and later professor, when they appointed a professor, as there wasn’t a professorship associated with the department originally. They had a sizeable department of punch-card operators and at some later date and for most of the Caerphilly data we would pay them to enter the data for us. In the early days it would mean putting data onto punch cards and later
we used direct entry, putting them onto files on the computer and then later on the floppy discs. People like Janie were expert at it, she was very, very fast. I got to the stage where I was quite fast on these machines, I could cope with probably 7,000 or 8,000 key depressions an hour. Janie would probably be 50 per cent quicker and the ones who were doing it all the time probably double that speed. A lot of data could be punched in a short period of time.

AN: And the data you would then work on would be from your counts, you would have this on cross-tabbed tables.

PS: The counter-sorter could do cross-tabulations between variables. You could also make cross-tabs between three variables. If you imagine that you have got three variables, each with just three values each, then you got 27 different possibilities, which means you have got 27 different piles of punch cards. If another variable is added with four values, you are beginning to get yourself in one hell of a mess. Fred Moore was the obsessive one who could do that, because it really required somebody very orderly, very meticulous, to do it. Cross-tabs were about as far as you could go on a counter-sorter.

AN: I am just trying to get the feel of how you handled the data. So they took all the pro formas in the field, you would punch out a card and see if you could get your cross-tabs.

PS: Then you would do the analysis – electromechanical calculator, pencil and paper. And it would be up as far as doing multiple linear regressions with two or three variables. If you look at the papers from that date, that's very much the limit, because that was the limit of what was doable. Correlation coefficients weren't too bad if you had 50 subjects, but by the time you had 200 to enter and to calculate the correlation coefficient, my memory is that on the electromechanical calculator you could put the ‘x’ value down the left-hand end of the keyboard, the right-hand value down the right-hand end and if you multiplied that number by itself, looking at it as a whole number, you could get on the moving carriage the sum of squares of ‘x’ at one end, the sum of squares of ‘y’ at the other end and twice the cross product in the middle. Provided the numbers weren’t too large and provided there weren’t too many numbers, because eventually they would start overlapping each other. There was enough distance separating the numbers, providing the numbers weren’t too large or too many. But by the time that you get to 200, 300 or 400 numbers, you not only had the problem of overlap on the carriage, but to enter all those numbers correctly, is a tedious task to put it mildly. Again one has to be a bit of an obsessive really.

AN: So most of the things that you have been describing, the limit to what you could do when you first started was very much the technological rather than the methods not being described or appreciated.

PS: Well I think the two things in a sense go hand in hand, because certainly there would be a big theoretical literature on multi-variant analyses even then, but I think inevitably there’s a feedback. It must be to some extent, not pointless, but it's difficult to raise much enthusiasm one would think for producing methods that nobody could practically do. And I think once the means of doing them becomes practical, then I think the whole area expands, I think, in terms of the methodology. And probably it isn’t until you start using it that you find the problems associated with the methodology, so that the two things work in a circle really as to what’s practical. Now we can do this thing, is there a better way? Is there a better theory for what we ought to be doing? So I think it just goes round and round really.

AN: Well the things I was thinking of was, for example, the things you see in Richard Doll and Austin Bradford Hill’s first papers on smoking, when they gave some graphs they were talking about, they weren’t using ratios, there wasn’t an appreciation that the ratio could calculated was a useful statistic. And yet there was no technological thing to stop them.

PS: Well I think the theory was there but in a sense it wasn’t used, in a lot of cases it wouldn’t have been used because you simply couldn’t. It would just have been so time-consuming, because as soon as you get round to, you know with the multiple-logistic regression you have to reiterate, in other words you start off with
some estimates and then you do the calculations and then you get some more estimates and then you go round and you do it all again. Whereas with the multiple linear regression, OK you just have to sort the simultaneous equations once and that’s your answer. That was enough of a problem, but to have to go round and do it again and again and again and again, the chance of getting it correct by the time you had finished it, was, well the thought of then setting out and doing it all over again to see if you got the same answer you know, I think one would just have been doing these things forever. I may be wrong, and maybe if you look in the literature you can find one that was done by hand, but I suspect that you probably can’t, not one that’s of any size anyway.

AN: I am just interested in how much, given that technology was such an important part of this, how technology developed in the Unit and what links the Unit had with the computing and statistics side of things to take this forward.

PS: Well the link was always with the university computing centre. The first computer of any size, which I think certainly the undergraduates are still not permitted anywhere near, which at least provided a service and a number of people could use, the Stantec Zebra, which I told you about before, I think were housed in the physics department and literally it was a valve machine and it was a one-person at a time job. The ICL 4 something or other I can’t remember, they went through a private 450 I am trying to think whether it would allow multiple users right from the beginning, certainly our data entry and our programme entry was always by punch card initially. Whether it had terminals attached right at the beginning, I suspect probably not and how long afterwards those appeared, when we managed to get a terminal attachment, I find it very difficult to date. We had one in the Unit, a teddy-type turbo, literally a real noisy contraption. It lived down in the basement and sat on the main sewer cover, it wasn’t the most pleasant of places to sit and work, and to be fair to date that I am sure that it was Selwyn St Leger who was the keenest enthusiast for having this, so it’s certainly arrived after the date of Selwyn’s arrival [in 1974]. As for the university machine, I cannot remember the time at which packages of any sort first appeared. I don’t know. I would have thought late seventies, yes, it must be by then. I don’t know. I know the ones I used. The first one I used was a thing called the Rothamsted General Survey Program, which was again essentially a tabulation thing, basically produced tables, you could make it produce really very pretty tables, ones that you could, literally, if you wanted to, you could have produced them so that if you had a decent photocopier available in those days, you could have stuck the thing straight in a paper in fact, because they were that good, you really could. They were far, far superior to anything you ever got out of much later versions of things like mini-DANS and SPO-sets. Vastly superior. But that thing you had literally a three-stage program. The first one was a program written in the Rothamsted General Survey language to describe the data, tell it that certain things were variable, or there was a factor with five levels or whatever. The second part was a FORTRAN program, which physically read the data in, and the third part was in the tabulation procedure, which produced the tables, you could label things and I am trying to remember what operations you could do. I don’t know whether we were producing tables, probably not terribly much. Because I see a lot of system 4s, a lot of universities and a lot of research places, those were the machines that they had, they basically vied with this, was at the time when all the computer companies were merging, and I think it was originally an English Electric machine, because ICL’s main range of computers was called the 1900 series in those days and I think the ones with 450, 470, 490, I think these were English Electric. They had merged or ICL had taken them over, and presumably Rothamsted had one of these machines, because this package was originally written almost directly for that. But the university was interesting really, the computing set up, because it was one of the very first networks of computers anywhere, because there was a network between what was then, the one we used was University College Cardiff, in those days, UWIST [University of Wales Institute of Science and Technology] also had a system-4 machine, Bristol had one, Exeter had one, and I rather think Bath had one as well, and again dating this I can’t remember, it’s a struggle. They were linked together in a network, and when I used to run, for example, the Rothamsted General Survey package, the job used to run in Bristol, because that’s where that piece of software was. One of the main things was they didn’t have huge amounts of back-up devices, so in order to avoid the different machines all keeping the same software, certain bits of software were kept on this machine and other bits on somewhere else.
I used Exeter’s machine, this must be late seventies, even early eighties, for a whole series of taped jobs that we did, because one of the huge jobs that we did, starting in the early seventies, was cervical cytology. Hubert Campbell and Archie, with the Department of Health I think, had been the originators of the study. I think in the early 1960s the plan had been to have three or four areas of the country, which had a cervical cytology screening service, and the rest of the country wouldn’t have one, in the hope that they would be able to see whether the damn thing worked in terms of reducing mortality rates. No doubt at the point in time at which that was originally set up that was a reasonable thing to try and do, basically on the back of the experience from British Columbia, which was the only place that had any screening programme and were saying that it worked and people were asking what’s the evidence that it works. And in those days, our notions about how you would try and decide whether a screening programme did produce some benefit in terms of the natural history of the disease, they are very much more sophisticated now than they were in those days. But as it turned out the notion of having three or four areas with a screening service and the rest of the country without, so that you could compare the mortality rates in those with and those without, as it turned out was doomed to failure for two reasons I think. First of all because the areas chosen probably weren’t big enough, nobody had any terribly clear idea over what time period you would expect the effect to take place, did it happen within five years, ten years, twenty years? And I don’t think anybody very clearly knew this, and almost certainly the answer is that it is at the end of that range rather than at the beginning of that range and certainly by the early 1970s everybody either had a screening service, or was clamouring for one, shouldn’t be left out, etc. etc. So we really rather got landed with the job of trying to figure out what we could get out of this vast mass of data, because they were screening, by the seventies, they were screening 20,000-odd women a year, smear tests, and cards, that was the number we were doing. I remember there was a large meeting with [Professor] George Knox who was the expert on the subject and one of the things they thought they would like to try and get out of this was some measure of the incidence of what was then called carcinoma in situ. There was still at this stage, there seemed to be a huge discrepancy between the prevalence of carcinoma in situ, certainly in the incidence of invasive cancer, there seemed to be vastly more of the former than there were of the latter. And one of the things to try and do was to try and get an estimate of the incidence of carcinoma in situ, which is a difficult problem because you don’t just need really two examinations, because of the false negative problem – you really need at least three examinations.

AN: So the first two would define whether they were normal cells and the third would show that they hadn’t changed?

PS: Well, let me think which way round it is going to be… Obviously you start off with those, you are interested in those who are negative to start with and what you are interested in is what proportion of those who are negative at one point in time become positive at a second point in time. But because of the false negative problem, some of those who are apparently negatives to start are going to be truly positive anyway. So some of those that go from negative to positive haven’t necessarily gone from negative to positive, they have gone from an unrealized positive to an observed positive. So they haven’t changed, it’s the method measure. Basically you have really got three factors, you have got prevalence, incidence, and a false negative error rate, and you need three equations really to solve these things and you really needed three examinations. But the biggest problem initially was, how the hell do we handle what was now I don’t know, I guess records for a 100,000 smear tests, something like this. That was the first problem and the second problem, I think it was George Knox who pointed it out, well you are not going to be able to get any idea of incidence, because you haven’t dated it, all you have done is put the year of the test and therefore if you got one test with 1971 and another test with the year 1972 on it, the interval between them can be one day or two years. And you have absolutely no idea. So I think it was for about 100,000 smear tests, we then had to go back to the records and make out little forms to put the full date in and myself and a couple of others, I can’t remember who the devil they were now, did a lot of this by candlelight during the miners’ strike of 1974, during a week or ten days, we got stuck in. Because the cervical cytology unit was at 30 Richmond Road, just up the road. But we used to handle, I did this with a chap called Peter Samuels, who was a lecturer in medical computing in Hubert Campbell’s department up in the Heath [Hospital], and we worked on, the MRC had its own mainframe computer in Pentonville Road [London], its own computer unit, and the university machine at that time wasn’t prepared to try and handle jobs of tens of thousands of punch cards, this really would clutter the place up too much. So we used to go up to London and run this machine
ourselves at night, because they wouldn’t do the job during the day either, because it just blocked the whole dam machine up, so that we would ring up in the morning and say is the machine going, and if, ‘yes it is’ we would get on the train and up we would go. It was basically reading the cards in tens of thousands, through the card-punch reader, checking them, producing error messages, sticking them onto tapes, and then later creating records for individuals, which contained both the smear data and the clinical data as well, and then later trying to get something out of it. I think if you look back at the net result of what it produced, it didn’t produce terribly much for the amount of effort that went into it. Although it taught me one hell of a lot about computing to be fair to it.

AN: And you said that at the beginning when you came data would tend to land on your desk with ‘now analyse that’. Can you remember the first study when you were involved in the design stage and how your role changed over time?

PS: I can’t put a date on when statisticians became involved in the design stage of projects. I think things slowly developed. Certainly when Peter (Elwood) took over the Unit in 1974, he introduced a weekly staff meeting at which somebody would give a presentation, either of some analyses or some results, or a proposed piece of work etc., so that to some extent everybody got a little bit involved in the design of these things. I think over time we found that people tended to get a little bit overly critical and then others shied away from presenting things, because they didn’t want to get cut to bits, which tended to happen. And probably after that, it then later evolved more into the people who were involved with a particular proposed project, literally sitting down and discussing it. But I suppose what also happened as well was that the size of the projects got larger and larger. There were some exceptions, even in the 1960s, some of Peter’s anaemia studies are quite large numbers of individuals, but some of those were never even put on punch cards. The smaller ones were, but there were one or two that I think weren’t. They might have been on edge punch cards. It’s very difficult, I think to date changes in things, because the changes were usually gradual, rather than incremental, although certainly there was an incremental change, I think when Peter took over. He worked very hard at trying to decide how he was going to manage this unit and he talked a lot with a chap called Tony Johnson who was a lab manager from down in west Wales, who was very interested in the theory of management of organizations, and he introduced, in particular, these weekly staff meetings. We had staff meetings under Archie, but my memory is that they were much less regular and they were much more to do with routine things. There must occasionally have been presentations of planned bits of research and also results, but my memory is very hazy about these things. Much less so for things like the arrival of the Monroe Epic 190 and what you could and couldn’t do with it, than for the way the Unit was managed. In a sense, it’s probably partly me, but it’s also partly the statistician’s role, in that he didn’t really manage the research, that was essentially the epidemiologist’s job, and the statistician was there as adviser and essentially a calculating machine.

AN: During your career did you initiate pieces of work or was it always in tandem?

PS: My work was almost always in tandem, particularly once Caerphilly got up and running and I got involved with that, but even before that, there was always more than enough to do. In the first few years, there certainly wouldn’t have been from within the Unit, which is why Hubert Campbell got me involved with the gynaecologists at the Heath Hospital and others as well. But after that there were Peter Elwood’s aspirin trials for which I did the analysis. There was always cervical cytology, which in one sense was entirely my pigeon. It wasn’t really an Epidemiology Unit project, although Archie was very interested, Peter and Michael Burr didn’t want to know. I was interested in learning to handle these large volumes of data and learning how to use the computer. It was very useful from that end, but it was certainly extremely time-consuming. Then once the aspirin trials finished, I guess the second one, I think The Lancet paper for the second one is early eighties. I wasn’t involved in Caerphilly immediately, because I think that Selwyn St Leger was the official statistician to the Caerphilly study when it initially started up. John Yarnell had done a variety of studies in Caerphilly, and this was essentially just another one. His initial study was the wide-age range study of 700-old men aged 30 to 69 years. My memory is that it was a unit review in the early 1980s, which said, ‘Well, that’s an interesting piece of work, but it’s nothing like big enough, make it larger’. Hence it was expanded into the size that it became, 2,500. It was about then that I probably started getting involved,
I guess partly because Peter's aspirin work had finished. I think the paper on the second aspirin trial in The Lancet was some two-thirds of the way through the first phase in Caerphilly. I think that would be about the date of my first involvement with the study. My involvement with aspirin finished about this stage, and by then the cervical cytology had outrun its course. I think we had got about as much out of it as we thought we were ever going to.

AN: You mentioned that there was a criticism of the size of the Caerphilly wide-age range study, and in retrospect one can look back on some of the studies that the Unit did, methodologically one wouldn't be able to criticize them, but one might say that it would have been nice if they were slightly bigger. Was that something that you were aware of? Or is it easy with retrospect for someone to say?

PS: I think it's probably easier in retrospect to say. The criticism could certainly be levelled at the Caerphilly study, that it was not big enough, and that's again like most things it's a combination of factors in that, you know I wouldn't say that it grew like Topsy, but what it ended up with is not exactly what it was planned to be originally. John's wide-age range study had a specific set of objectives, and probably the size of the study was large enough to meet those objectives, but I think what the review committee were saying was, 'well OK it meets those objectives, but those aren't really the interesting objectives'. The interesting objectives would be to have a look at how these things relate to incidence of disease, and you haven't got anything like enough to do that. And that's really what you ought to be doing. And certainly I can remember then sitting down and trying to do calculations on what numbers do you need if you want to be able to say with a given power that this difference in fibrinogen level is significantly associated with incidence of ischaemic heart disease. And as always with these things, the difficulties are that depending on which variable you take, whether you take fibrinogen or total cholesterol or HDL (high-density lipoprotein) or something else, they all provide different answers to that question and what gets much more difficult from the numbers point of view is you have very little idea when you start off doing a thing like that what the interrelations are between the various variables and therefore how you would allow for that in the calculation of the numbers is difficult if not impossible. And the choice of the numbers comes down eventually to a compromise between what do we think we need, for what we think at this moment in time are our principal objectives, what can we handle, and what's available. In Caerphilly, without spreading the net geographically very much wider, we took literally all the men aged 45 to 59, it was all of them in Caerphilly, and a few outlying areas, it took all that was available. That was about what we thought we could handle, and which was roughly what we thought we might need to meet what we thought were the principal objectives. The problem is that the objectives change over time, everything changes over time, and all of a sudden you find that you haven't really got necessarily exactly as many as you would like. The same to some extent, I think, could be said of other studies I guess.

First, the aspirin study was expanded greatly and in a hurry because of Hershel Jick's findings. Because of Jick's work the study ended up with, I think, 1,600 or 1,700 individuals. Whether it would really have been much better if it had been 3,000 or 4,000 is a moot point, because if you had looked at aspirin and ischaemic heart disease, say, in the late 1970s, I think you would have found something like five or six sizeable randomized controlled trials, all of which (bar one) showed really quite similar sorts of results. Typically there's about a 25 per cent benefit. None of these trials, by and large, were statistically significant in themselves, and one, the AMIS (Aspirin Myocardial Infarction Study), which showed next to nothing, was by far the biggest of the lot. I have always felt that I would much rather have five separate, disparate trials of 10,000 individuals rather than one of 50,000. Taking the aspirin example, if AMIS had been the only one that was done, the big one, aspirin might have disappeared forever from the face of the medical earth, in terms of thrombotic disease. There's a great advantage to having large samples, but if it came down to a choice between one large and rather more medium ones, I would go for the rather more medium-sized ones every time, with different people doing it, slightly different designs or different subjects. I think it is much safer.

AN: Was there a sense that the effect sizes that you were interested in detecting were bigger than perhaps the sizes that Sir Richard Peto and others now would be setting up their trials to detect?
PS: Well I think that’s, you see the clinical trials are rather different from something like aetiology I suppose, or whatever one would call the Caerphilly, I think in the one case you are asking, I suppose ultimately the question you ask, is ‘how big do we need this trial to be’, to give us the given power of getting the statistics significant, a difference that we view as being clinically worthwhile. Although I am sure you can get plenty of disagreements about what’s clinically worthwhile, you can get roughly a broad consensus surely, whereas in the aetiological thing the difference you are talking about is you don’t have any worthwhile equivalent clinically really. Are we looking for a difference in fibrinogen of 0.5 [gram] a litre, 0.2 of a gram of a litre or whatever, and you don’t have quite the same constraint, constraint isn’t the right word, but you don’t have the same method of choice I don’t think as to what’s sensible to choose. I suppose aetologically what’s sensible to choose is the difference, which you thought you might be able to change by some reasonably simple interventions at a later date. But when you say that you see, if you were then sat there doing these calculations in the 1980s in terms of fibrinogen, you didn’t have the faintest idea in the 1980s, and let’s face it we don’t have them now, what simple intervention you can have to change fibrinogen levels other than to get people to give up smoking. You don’t have a means of setting your levels for an aetiological study in the same way as you do, imprecise as they may be, and controversial as they might be, in a clinical trial, where it is much easier I think to choose a number. You can get people to disagree with it no doubt, but it’s much easier to choose a number.

AN: Caerphilly’s quite interesting because I think it is the first real conscious launch into setting up a large cohort study. That would not be entirely true because the Rhondda studies and the follow-ups and so on, the Unit in a sense came late to prospective studies. I was just wondering again why you thought that was or whether you were a party to discussions.

PS: Well Archie had always followed up most of his studies. I suppose what’s different about Caerphilly really is that the individuals are examined repeatedly, rather than the follow-up being purely for an event of one sort or another. The first time I . . . somebody eventually came in and said to me one day ‘you must come out and see what’s done in the field’, because I had never seen this before. And what they took me to see was the follow-up in Staveley [Cumbria] in fact. That’s what it was, because I have a very vivid memory of that because it was done in a church hall and the church was next door, and when you went into the church hall there were four desks, behind which were sat Archie Cochrane, Ian Higgins, an American called Ben Ferris, and somebody else, and these were the four interviewers and as was certainly Archie’s wont in those days, everything was done to try and eliminate the bias, and the men coming in were randomly allocated to one of the interviewers, and if you try and explain to these gentlemen why they should be sitting waiting – this is very topical this subject of waiting lists – there would be three or four or five sitting waiting, when at the four desks of the interviewers there was one doing some interviewing and the other three were sitting there twiddling their thumbs, because the next three or four were all allocated to see the one interviewer who had currently got somebody sat in front of him, because nobody had thought to do a blocked randomization. So Hubert Campbell and I decided, or Hubert Campbell decided, we would have to block this randomization, because this was crazy, word would get round that you were being made to wait unnecessarily, and I have distinct remembrance of sitting in the church next door with Hubert Campbell, getting the hymn book, turning it upside down, and opening and saying right the one on the right-hand side last digit is number seven and we did a block randomization doing this using the church hymn book. So I knew prospective studies of a sort were really always there. You would call those cohort studies, I think you would, but it is fair to say that Caerphilly is the first venture into a large-scale cohort study and why not earlier? Again I would guess a combination of things, presumably none of the studies that anybody wanted to do earlier, anybody had found any great necessity for doing a large, long-term cohort study, and they do take a lot of staff that’s the other thing. If you look in the 1970s and the 1980s, well even in the 1960s the place was much smaller, by the time you get to the 1970s it had expanded a bit, but in a sense all the epidemiologists had their own bits of work, for example, so that Michael Burr had his own bits of work, John Yarnell had his own bit of work, Estlin Waters, who had gone by then, I suppose he had his own bits of work, and Peter had his. They were all different, it wasn’t really until we were coalesced into the same subject matter that it was really a practical possibility in terms of personnel. Just how conscious the decision was to do that, you see I don’t think anybody sat down and said ‘let’s do a large-scale cohort study of ischaemic
heart disease’, it did in a sense grow a bit like Topsy. It started off as John Yarnell’s wide-age range study, and then it became really with the review body saying well what you really ought to be doing is make it bigger. So okay we ended up with a cohort of 2,500. And then again with Peter’s interest, it was still very much John Yarnell’s pigeon at that stage, because Peter was still very heavily involved in aspirin, but with his interest in aspirin and his long-term interest in platelet function, then all of a sudden you now get the possibility that you can do platelet function tests in the field, with Renaud’s mobile laboratory. Okay right, we can examine the men again because there are all sorts of other things we can do like platelet function, and measure lipoproteins and white cell differential etc, etc, so we would examine them again and do it again. And then once you have examined, it develops a momentum of its own then, because once you have done the tests, you have really got to examine them, well you have certainly got to follow them up again in order to get outcome measures. So that I am not so sure that there was ever. I don’t think anybody sat down in the 1980s and said ‘let’s do a 15–20-year cohort study of 2,500 men in Caerphilly’. I really don’t think that ever happened.

**AN:** We were talking a little bit about cohort studies and one thing you mentioned that I wanted to pick up on was the issue of the size of the Unit and the resources. I was wondering how much the work was in part determined by the size, which seemed to be constrained and designed by MRC to some extent.

**PS:** It was always constrained by that, because I can’t put an accurate date on this, but for most of the time that I was with the Unit the MRC didn’t permit its units to apply for grants from other bodies, presumably on the basis that it would imply that they weren’t funding their own units properly. And therefore you had the set of staff, you had the complement of staff that were decided at the five-yearly reviews and that was that. You could try at those five-yearly reviews to ask for more and you could I think, and I am sure that basically Archie no doubt, and Peter certainly I think, would on occasion have asked for half a person or a person here or there for a particular study which was in progress. But you really couldn’t just suddenly say ‘well we want to do this and we need five extra staff to do it’, that simply wasn’t possible. The point in time at which MRC did then later permit their units to apply for grants from outside bodies, again I can’t date accurately, but I would think it comes considerably later, certainly not at the beginning of Caerphilly. I would guess it as the middle or late 1980s, but how far that is right, but that would be my guess. So, yes, you were constrained by what staff you had and so in terms of something like Caerphilly, each of the epidemiologists tended to have their own staff to some extent, they tended to have their own fieldworkers and their own assistants. That’s not entirely true, some would be shared across people really, but I suppose in the 1980s we had got three epidemiologists who were there all the time, John Yarnell, Mike Burr and Peter. With the finish of the aspirin work, and there are other areas that I haven’t touched on like the industrial stuff, the cotton workers and the flax workers, although those were not major, although Bolton and Oldham were, but the others weren’t major users of fieldworker time, but with the end of the aspirin studies, then Peter’s fieldworkers were by and large moved on to Caerphilly. They were already doing it anyway, because they were partly the ones that John was sharing. But you see Michael never got involved really in the cohort study of Caerphilly. He ran a lot of parallel clinical trials of all sorts of things from vitamin C deficiency from small to really very large, but he had his own staff that were doing those, who were then therefore, by and large inevitably unavailable for other things. I guess Peter as director, if he had wanted to, could have said ‘right we want to make Caerphilly twice the size it is and therefore everybody will have to work on it’, but I think certainly Michael would have said ‘well yes I will go and work somewhere else, because I can’t do what I want to do now’. And so it has to be constrained by the resources.

**AN:** And you told us a bit about the way each epidemiologist would have their fieldworker. How would statistician time be apportioned round the Unit, how many of you were there and how would it work?

**PS:** Well we were divided out, the statisticians. I am trying to think what the maximum number of us there ever were at any one point in time. You see when I arrived Hubert Campbell was there half time and he was nominal, but to be fair he worked much more in the medical school than he did in MRC. Howard Lovell who originally worked with Bill Miall in Jamaica came and worked for about a year with us I think before...
he went to Aberdeen. And Margaret Abernethy was there when I started as well, but she departed, then David Hole came and he worked for I suppose, I don’t know how long David worked for, four or five years, I am bound to forget somebody in this. Then Selwyn St Leger, then Janet Milbank, Janet Pickering. Where did Janet fit into this? I think Janet probably fitted in after David Hole and by and large before Selwyn St Leger, although I think they must have overlapped. Although my memory suggests that they both had the same room at the top of the stairs on the left, in which case maybe they didn’t overlap. So Janet, Selwyn, Barbara Butland.

AN:  Ruth Holliday?

PS:  Not a statistician you see, no, no computer programmer. She did odd bits of statistical analysis, but that was not what she was. She was a programmer. So Barbara Butland, and then Liz Lim really. It wasn’t until David Poor appeared, again God when? Middle eighties I suppose that we had a data-entry man, well not a data-entry man, but a data-control man, I suppose, that was really his job, so he was a computing man, not a statistician. Ruth came to us basically because of the closure of a unit somewhere else. And she worked mainly with Michael but not entirely with Michael, but she did a lot of the work on particularly his fat, fish, and fibre study. But she was an extremely good programmer. She also did work for Caerphilly, she wrote the programs for the organizers, written in Basic for the organization of the control function data, she made the programs. The psychologist who produced the programs which ran the tests on the computer, she first of all made them user friendly and usable for people in the field, produced the backup systems, the archiving systems for them, and the analysis programme.

AN:  Did she write the analysis program?

PS:  God I don’t know. But anyway she was a good programmer, but not a statistician. But she did odd little bits of statistical analysis.

AN:  It sounds like there would have been two of you.

PS:  Usually there would be two, there may have been the odd occasion when there would be three, but that would be pretty unusual and there were also occasions when there was only one, and that would have been me. And we were just divided around between whoever wanted things and to some extent it was the order of priority was he with the greatest clout in terms of who wanted their stuff. And that was basically the way it worked. Certainly I tended over the last 15 years to work almost entirely on Caerphilly and Speedwell. I did odd little bits for Michael, but probably not terribly much, most of it would have been done by Ruth Holliday and also with Liz Lim of course. Liz did a lot of Michael’s allergy stuff. So yes Liz would have done most of his stuff. She also did bits of Caerphilly as well too, as did all the others.

AN:  You mentioned that you did your MSc and took a while to complete it. I don’t think you went on to do a PhD did you?

PS:  No.

AN:  Did you ever consider that?

PS:  No, never really thought about it. Again there are probably three constraints, the main one probably being me, I was never entirely sure that, I couldn’t really figure out what to do it on. I became very much a practical statistician, doing things. My theoretical bent was not very good I don’t think, so a) I guess I wasn’t really at all sure that I was capable of doing one, b) it wasn’t easy to see, there wasn’t any obvious advantage to doing one, and c) I couldn’t see where the devil I would get the time to do it from. Doing it, and doing it in a sense on the job almost didn’t seem possible really. Because frankly there was always more work to be done than there were statisticians to do it. There was always a pile of stuff waiting to be done. And that’s inevitable once a big cohort study like Caerphilly gets going, because there are all sorts of people with all sorts of ideas as to what use you might put the data to and it is interesting I think with those sorts of studies
it seems to me that you require almost different mixes of personnel at different points in time along the study, in that you need initially lots of fieldworkers, or what we would call fieldworkers anyway, and the statisticians you could do with relatively few of them, but by the end you need the opposite mix. It’s not easy, if done within the MRC unit type place it was impossible to change, you couldn’t sack a couple of fieldworkers and get another statistician, you would have had to sack a couple at least, and that wasn’t the way that things worked. And the other thing is that because there are fieldworkers and because they have to be kept busy, then you keep on collecting data, and therefore the people organizing the data and trying to analyse the stuff are permanently running to try and catch up in fact and they never do. And it’s built into the system somehow.

**AN:** But did you feel like, although there would only be a couple of you at any one time, did you feel quite isolated as you describe yourself as a practical statistician, or were there strong links with the university and other people?

**PS:** Statistical links, no there probably weren’t. And again this is probably me rather than the job. In part the university department of statistics was very much a theoretical department and therefore I had much more in common with the department of medical statistics and computing up at the Heath, but when Hubert Campbell departed, when he retired, they didn’t appoint another professor, they appointed a senior lecturer who was Ted Coles and he wasn’t a statistician. The best way I can put it is that the department’s name changed, to the Department of Medical Computing and Statistics and that change is significant in that the statistical side, I won’t say went downhill, but the computing side was given a very much greater emphasis and obviously it was needed in the medical school at that point in time. But there weren’t really, I am trying to think how many statisticians there were there, probably only Robert Newcombe for a long period of time, and he had the whole medical school to look after, so he was up to his ears in work. I was up to my ears as well, and somehow one almost begrudged the time, probably entirely wrongly one suspects, in spending the time making links with other people and yes, I think undoubtedly it would have worked better if we had been within a much larger organization somehow. I think there is undoubtedly a critical mass for these things, below which it shouldn’t be allowed to fall.

**AN:** I was just wondering how you kept up to date or exchanged tips with people on how you do this and do that.

**PS:** Well in a sense I am not entirely sure that I did you see, in that one found oneself at the end still doing multiple logistic regression analysis, which one had been doing as soon as there was a computer around that enabled you to do it. I guess one does change, but I suspect that if you look at the methodologies that are used in the papers on Caerphilly from the last five years, they wouldn’t be very much different from the ones in the first five years. And whether they should be or not is a different question I think. Whether one used in the papers on Caerphilly in common with the department of medical statistics and computing up at the Heath, but when Hubert Campbell departed, when he retired, they didn’t appoint another professor, they appointed a senior lecturer who was Ted Coles and he wasn’t a statistician. The best way I can put it is that the department’s name changed, to the Department of Medical Computing and Statistics and that change is significant in that the statistical side, I won’t say went downhill, but the computing side was given a very much greater emphasis and obviously it was needed in the medical school at that point in time. But there weren’t really, I am trying to think how many statisticians there were there, probably only Robert Newcombe for a long period of time, and he had the whole medical school to look after, so he was up to his ears in work. I was up to my ears as well, and somehow one almost begrudged the time, probably entirely wrongly one suspects, in spending the time making links with other people and yes, I think undoubtedly it would have worked better if we had been within a much larger organization somehow. I think there is undoubtedly a critical mass for these things, below which it shouldn’t be allowed to fall.

**AN:** I was just wondering how you kept up to date or exchanged tips with people on how you do this and do that.

**PS:** Well in a sense I am not entirely sure that I did you see, in that one found oneself at the end still doing multiple logistic regression analysis, which one had been doing as soon as there was a computer around that enabled you to do it. I guess one does change, but I suspect that if you look at the methodologies that are used in the papers on Caerphilly from the last five years, they wouldn’t be very much different from the ones in the first five years. And whether they should be or not is a different question I think. Whether one used in the papers on Caerphilly in common with the department of medical statistics and computing up at the Heath, but when Hubert Campbell departed, when he retired, they didn’t appoint another professor, they appointed a senior lecturer who was Ted Coles and he wasn’t a statistician. The best way I can put it is that the department’s name changed, to the Department of Medical Computing and Statistics and that change is significant in that the statistical side, I won’t say went downhill, but the computing side was given a very much greater emphasis and obviously it was needed in the medical school at that point in time. But there weren’t really, I am trying to think how many statisticians there were there, probably only Robert Newcombe for a long period of time, and he had the whole medical school to look after, so he was up to his ears in work. I was up to my ears as well, and somehow one almost begrudged the time, probably entirely wrongly one suspects, in spending the time making links with other people and yes, I think undoubtedly it would have worked better if we had been within a much larger organization somehow. I think there is undoubtedly a critical mass for these things, below which it shouldn’t be allowed to fall.

**AN:** I was just wondering how you kept up to date or exchanged tips with people on how you do this and do that.

**PS:** Well in a sense I am not entirely sure that I did you see, in that one found oneself at the end still doing multiple logistic regression analysis, which one had been doing as soon as there was a computer around that enabled you to do it. I guess one does change, but I suspect that if you look at the methodologies that are used in the papers on Caerphilly from the last five years, they wouldn’t be very much different from the ones in the first five years. And whether they should be or not is a different question I think. Whether one used in the papers on Caerphilly in common with the department of medical statistics and computing up at the Heath, but when Hubert Campbell departed, when he retired, they didn’t appoint another professor, they appointed a senior lecturer who was Ted Coles and he wasn’t a statistician. The best way I can put it is that the department’s name changed, to the Department of Medical Computing and Statistics and that change is significant in that the statistical side, I won’t say went downhill, but the computing side was given a very much greater emphasis and obviously it was needed in the medical school at that point in time. But there weren’t really, I am trying to think how many statisticians there were there, probably only Robert Newcombe for a long period of time, and he had the whole medical school to look after, so he was up to his ears in work. I was up to my ears as well, and somehow one almost begrudged the time, probably entirely wrongly one suspects, in spending the time making links with other people and yes, I think undoubtedly it would have worked better if we had been within a much larger organization somehow. I think there is undoubtedly a critical mass for these things, below which it shouldn’t be allowed to fall.

**AN:** I was just wondering how you kept up to date or exchanged tips with people on how you do this and do that.

**PS:** Well in a sense I am not entirely sure that I did you see, in that one found oneself at the end still doing multiple logistic regression analysis, which one had been doing as soon as there was a computer around that enabled you to do it. I guess one does change, but I suspect that if you look at the methodologies that are used in the papers on Caerphilly from the last five years, they wouldn’t be very much different from the ones in the first five years. And whether they should be or not is a different question I think. Whether one used in the papers on Caerphilly in common with the department of medical statistics and computing up at the Heath, but when Hubert Campbell departed, when he retired, they didn’t appoint another professor, they appointed a senior lecturer who was Ted Coles and he wasn’t a statistician. The best way I can put it is that the department’s name changed, to the Department of Medical Computing and Statistics and that change is significant in that the statistical side, I won’t say went downhill, but the computing side was given a very much greater emphasis and obviously it was needed in the medical school at that point in time. But there weren’t really, I am trying to think how many statisticians there were there, probably only Robert Newcombe for a long period of time, and he had the whole medical school to look after, so he was up to his ears in work. I was up to my ears as well, and somehow one almost begrudged the time, probably entirely wrongly one suspects, in spending the time making links with other people and yes, I think undoubtedly it would have worked better if we had been within a much larger organization somehow. I think there is undoubtedly a critical mass for these things, below which it shouldn’t be allowed to fall.

**AN:** Most people that I hear talk say you do all these fancy maps and analyses and you should always do the basic stuff anyway.

**PS:** I think you must always do the basic simple stuff like ask yourself what’s the mean, what’s the distribution look like, what’s the reproducibility of these things? That’s one of my slight concerns with computers, that it is a little bit too much like shoving it in a black box and answers come out at the other end and you don’t get your hands dirty and see the data, whereas I think we used to write the damn stuff down in pencil on bits of paper, and to some extent I still do that and I think it’s critically important that statisticians really do have a very good look at the raw basic data to see physically what they are handling. And also to go back to the questionnaires to have a look to see, OK that’s the number but what does it really mean? Let’s go back and have a look at a few of these things. And until you really do that, I think it is very easy to assume that the data is better than it is. I don’t want to disparage the daily collection, because that in itself is an extremely difficult task, but it needs to be recognized that it’s far from perfect you know. And the trouble is once it’s a number on a computer, then it’s a little bit like it’s cast in tablets of stone and it shouldn’t be treated like
that, one should look at it all with a certain air of cynicism, no cynicism is the wrong word, scepticism, I think. Because with the best will in the world you know labs with all their fancy machinery these days, they still can and do turn out numbers which are little more than piles of random numbers. I remember David Bainton, this must be from his gallstone study, looking at a set of results one day and ringing up the lab and saying ‘well didn’t you see anything odd about these numbers?’ ‘No, we didn’t look, the numbers are the numbers and they are always alright’. ‘Well what about your quality control’ this was in the early days of quality control, and I said ‘oh we used to run those but we don’t do it any longer’, and these were serum bilirubin results that I remember. And David said ‘they would all have looked bright yellow the people I took that blood from if these were the numbers’ and you only have to look even at what you would assume to be the simplest tests, you know things like a total white cell count on a cell counter. And if you look at the Caerphilly stuff and you look at that over periods of time, and there are shifts in the average levels, which means that you know the stuff is drifting up or whatever, whether by change of operator, change of reagent, cleaner machines or whatever, and these things always have to be looked at. By and large most of these things are fairly irrelevant in terms of the overall project. They are too small to be terribly important, but on the other hand you will get things like, if you look at the Caerphilly antithrombin-3 results of which there are two in Phase one, one of them is totally and utterly useless, because when you look at it you will find there is an incremental step change in the results, and they introduced a new method, part way through, oh you will get better results with a new method, but they don’t think to tell you and this renders the whole thing completely useless. But if you just look at the results, even if you just look at a complete distribution of those antithrombin-3 results, the distribution probably looks a little bit odd, but no odder than distribution of a few other things. It’s only when you look at it over chronological times, that all of a sudden you find ‘hello something crazy has happened here’ and you have to do something with it, you either scrub the results, but unless you do the basic looking at it, and it’s quite a lot of looking, it isn’t just a simple casual look, but it is a lot of looking, it’s easy enough to do something which makes a complete nonsense of the stuff.

**AN:** Going off at a slight tangent, you stayed in the Unit for a long time, I was wondering if you had ever had itchy feet, wanted to move, and as you say many of your fellow statisticians came and went.

**PS:** Yes. The coming and going of the other statisticians, for most of the time the MRC had a system of three-year contracts for people, renewable, and they also would give permanent contracts before. And I don’t think that any of the other statisticians ever had a permanent contract, some departed because it was simply end of contract and MRC weren’t going to up the establishment of the Unit, which was two permanent statisticians I think, so there was that difference. As to why I never moved on, I guess this is the nature of the beast, this is the nature of me really. Two things I suppose really. First of all, for the vast majority I thoroughly enjoyed my time there, I didn’t have any great desire to go any place else. And also I suppose I didn’t have either the ambition, the imagination, call it what you will, to see if the grass might have been greener on the other side somewhere. Basically I guess really I enjoyed it. But although I am not sure that John Gallacher’s test would agree with this, I am not your thrusting type A individual, not really I don’t think. I was quite happy to potter along as a backroom boy doing all this stuff and I found it interesting and I enjoyed it.

**AN:** Another thing again, it may seem totally unconnected, talking to people about the Unit, I am quite struck by how often the old staff were often Welsh, but the first director was a Scot, the second director was an Ulsterman, it was very much an organization run from London. I was wondering how much this was seen as something that was relevant to the Welsh people in the sense that a lot of the studies were built on the commitment of the miners and their willingness to be studied and this was a natural place where these studies could be carried out, how much the unit responded to Welsh concerns, whether it had a Welshness to it or not.

**PS:** I have never thought about it like that. This is Julian Tudor Hart’s point essentially I guess that there ought to be, I don’t know what the right word is, there ought to be an interaction anyway between the research workers and the researched, in order to get something out of it. I think he is probably right that over the
vast majority of the time, the people who were researched didn’t get a lot out of it. A few will have done, but that would have been in a sense as an incidental, okay we found this and we are referring you to your GP. So that I think it wouldn’t have been looked upon as a two-way participation, and I think probably it would never have been looked upon like that. I may be wrong, and in a sense I am not the best person to ask, because I was never at the sharp end, meeting the people who were the subjects. In a sense they were all numbers on bits of paper to me, but the observation that the fieldworkers were Welsh and the scientific staff weren’t is not one that frankly had ever really occurred to me. The fieldworkers were Welsh because they knew and they lived in the locality. The scientific staff came from wherever they came from because they were the ones who applied for the jobs when they were advertised and they were the ones who were thought to be best for what was wanted. And I don’t think a question of nationality, I can’t imagine that it ever came into it.

AN: That’s good. It’s nice to hear that. It’s not that I am fishing that they should be.

PS: But as I say I don’t really think that the research was ever viewed as being done really to help the people.

AN: Peter, thank you very much for your time.

[END OF TRANSCRIPT]

Further related resources:

