



**From Bench to Bedside:
Tracing the Payback Forwards from
Basic or Early Clinical Research –
A Preliminary Exercise and Proposals for a Future Study**

Steve Hanney
Iain Frame
Jonathan Grant
Philip Green
Martin Buxton

HERG Research Report No. 31

August 2003

Health Economics Research Group,
Brunel University, Uxbridge, Middlesex UB8 3PH, UK

ISBN 1-902316-37-1

This report is available as a pdf file from: <http://www.brunel.ac.uk/depts/herg/pubs/internal.html>

Details of pre 2003 HERG reports can also be found at the above web address.

Copies of reports can be obtained by writing to:

Health Economics Research Group, Brunel University, Uxbridge, Middlesex UB8 3PH, UK.
Tel: +44 (0) 1895 203196; Fax: +44 (0) 1895 203330

TABLE OF CONTENTS

Preface	i
Contributors, acknowledgements and contacts	ii
Executive summary	iii
1 Introduction	1
1.1 Background	1
1.2 Potential large-scale study.....	3
1.3 Preliminary study	3
2 Bibliometric Analysis	6
2.1 Introduction.....	6
2.2 Methods.....	6
2.2.1 Citation tracing and mapping	6
2.2.2 Research levels.....	10
2.3 Results.....	11
2.3.1 Numbers of papers and self-citations.....	11
2.3.2 Half-lives	13
2.3.3 Research levels.....	14
2.3.4 Numbers of authors and addresses	16
2.3.5 Authors' addresses by country and distribution in the UK.....	17
2.3.6 Funding sources.....	18
2.4 Key Findings.....	19
2.4.1 Key observations about the body of work	19
2.4.2 Key lessons learnt about how to conduct such an exercise.....	20
3 Categorisation of Citations	21
3.1 Introduction.....	21
3.2 Literature review: categorising citations.....	21
3.2.1 The aims of analysing citations	22
3.2.2 Procedural issues: who classifies which citations?	23
3.2.3 What to include in a classification: motives and strength of impact	25
3.2.4 The location and frequency of citations within an article	28
3.2.5 Combining factors	29
3.3 Methods.....	29
3.3.1 Devising the template.....	29
3.3.2 The Pilot study	30
3.3.3 Application of the template.....	30
3.4 Results and lessons from the Pilot study.....	31
3.4.1 Importance of the cited article to the citing article	31
3.4.2 Categorising citations.....	32
3.4.3 Lessons from the pilots	34
3.5 Results of the expanded template application to all 2 nd generation articles ...	35
3.5.1 Overall importance of the cited article to the citing article.....	35
3.5.2 Importance of the cited article and location of the citation occasion	37
3.5.3 Overall importance of the cited article and classification citation	38
3.5.4 Article class and research level.....	40
3.5.5 The research setting	40
3.6 Key findings	41
3.6.1 Key observations about the payback	41
3.6.2 Key lessons learnt about how to assess payback.....	42

4 Qualitative Analysis	43
4.1 Introduction	43
4.2 Methods: original plan	44
4.2.1 Questionnaires to 1 st generation co-authors	44
4.2.2 Development of critical pathways by selected experts/peers	44
4.2.3 Interviews with researchers, peers and other stakeholders	45
4.3 Results of applying the methods.....	45
4.3.1 Questionnaires to 1 st generation co-authors	45
4.3.2 Development of critical pathways by selected experts/peers	46
4.3.3 Interviews with researchers, peers and other stakeholders	47
4.3.4 Analysis based on a selection of 10 key papers from Alberti	47
4.4 Findings from applying the four methods.....	50
4.4.1 Examples of important basic research	50
4.4.2 Basic research of varying significance	51
4.4.3 Research with major and rapid clinical impact	51
4.4.4 Early clinical research of mixed immediate impact but major long-term importance.....	53
4.4.5 Studies of commercial relevance	55
4.4.6 Studies that disproved an existing theory	57
4.4.7 Other studies on the selected list and the overall link to clinical practice.	57
4.4.8 Impact on career development and medical education.....	58
4.5 Key findings from the qualitative analysis.....	58
4.5.1 Key observations about the payback from the body of work.....	58
4.5.2 Key lessons learnt about how to assess payback.....	59
5 Lessons Learnt and the Way Forward	61
5.1 Lessons learnt about the exercise	61
5.2 From preliminary to large-scale study.....	63
5.3 Proposal for a large-scale study	64
5.4 Concluding comments	67
6 References	68
Appendices:	
Appendix 1: Funding sources for each paper	73
Appendix 2: Guidance notes for completion of the datasheet	74
Appendix 3: Template for categorising articles	75
Appendix 4: Template for categorising reviews	76
Appendix 5: Location and category of the citation occasions in five papers in the Essential group	77
Appendix 6: Questionnaire to original authors	78
Appendix 7: Critical pathway instructions	80

PREFACE

The Health Economics Research Group (HERG) at Brunel University is involved in a long-term programme analysing the benefits from health research, with an emphasis on the payback from health services research. This programme is being expanded, in collaboration with colleagues working on these issues elsewhere, to include exploration of methodologies for assessing the impact from basic or early clinical biomedical research. A central collaborator is Jonathan Grant, formerly Head of Policy at the Wellcome Trust, and now at RAND Europe. In simultaneously publishing two reports in the HERG Research Report series we bring together several elements of this research, and draw on them to make proposals for further work.

In HERG Research Report 30, *From Bedside to Bench: Comroe and Dripps Revisited*, Grant *et al* examine whether it is possible to replicate, and validate, the pioneering work of Comroe and Dripps in the 1970s. The latter traced back from then current clinical practice to the knowledge behind the advances. They claimed that more than half of the articles identified as making a key contribution to the clinical advances resulted from basic research. The attempted replication proved difficult, but Grant *et al* describe how they developed and applied an alternative methodology.

In HERG Research Report 31, *From Bench to Bedside: Tracing the Payback Forwards from Basic or Early Clinical Research--a Preliminary Exercise and Proposals for a Future Study*, Hanney *et al* describe a joint HERG/Wellcome Trust project that in part builds on the emerging findings from Grant *et al*'s study of the Comroe and Dripps methodology. Recognising the difficulties in tracing *backwards* from clinical practice, the project described here attempts instead to work *forwards* by tracing the impact from research conducted 20 years ago. Having described how the methods were applied in a preliminary study, the report goes on to outline how the work could be developed in a larger study.

The research undertaken for both reports was primarily funded by the R&D Directorate of the NHS Executive London, whose Director of R&D, Sally Davies, has been a stalwart supporter of such research and of its aim to provide an evidence-base for health research funding policies.

Martin Buxton

Health Economics Research Group

CONTRIBUTORS, ACKNOWLEDGEMENTS AND CONTACTS

The idea for this preliminary study was conceived by Jonathan Grant (Wellcome Trust) and Martin Buxton (HERG), who had overall responsibility for the project. Iain Frame (Wellcome Trust) and Steve Hanney (HERG) led the field work and analysis, with support from Philip Green (Wellcome Trust). The research team is most grateful to the NHS Executive, London Region, for funding the project.

Rhoda Nanre Nafziger and Karin Marley photocopied the research articles, entered the data and undertook the preliminary pilot study described in Chapter 3. Bryan Barratt was also part of that preliminary study. Tracey Young carried out the statistical analyses on the inter-assessor reliability data. Damian de Aspey designed and maintained the database, and Teri Jones later analysed some of the data. We are grateful for their assistance and for the important contributions to the work described in Chapter 4 made by Philip Home, Robert Tattersall, Moira Murphy, Sue Roberts and those of George (now Sir George) Alberti's co-authors from 1981 who responded to our questionnaire.

We are particularly grateful for the contribution of our Advisory Group: Sally Davies, Chris Henshall, Grant Lewison, Paul Lyons, Cathie Maciver, Mark Taylor and especially Philip Home whose contributions as an expert diabetologist were particularly valuable.

Finally, we are, of course, enormously grateful to George Alberti, whose assistance and willing co-operation were invaluable.

Corresponding author:

Steve Hanney
Research Fellow
Health Economics Research Group (HERG)
Brunel University
Uxbridge
Middlesex, UB8 3PH, UK
Email: stephen.hanney@brunel.ac.uk
Tel: +44 (0)1895 274000 ext 3709
Fax: +44 (0)1895 203330

To discuss plans for taking the study forward please contact Jonathan Grant (jgrant@rand.org) or Martin Buxton (martin.buxton@brunel.ac.uk)

EXECUTIVE SUMMARY

Chapter 1 : Introduction

- The members of the research team from HERG and the Wellcome Trust have conducted previous studies showing that it is possible both to assess the payback from applied health research, and to use bibliometrics to trace the links between generations of research and clinical guidelines. In another of the team's studies, however, it proved difficult to replicate the major study by Comroe and Dripps (1976) that had identified clinical advances and then worked *backwards* to show that they had relied on earlier basic research. Therefore, the study reported here sets out to use the methods developed in our previous studies of payback to undertake analysis that starts with more basic or early clinical research and traces the research lines *forwards* to clinical applications. Whilst this preliminary study involved preparation for a future large-scale study, it was hoped that it would also provide an interesting case study.
- Starting with the research outputs of one team 20 years ago, called the 1st generation papers, the preliminary study has three main elements: standard bibliometric analysis through several generations of papers; categorisation of the citations; and qualitative analysis using questionnaires, critical pathway analysis and interviews to trace the impact of the 1st generation of research.
- Diabetes and cardiology were suggested as possible topics on which to base the study. Initial reviews identified two bodies of research in diabetes as being potentially suitable for reasons such as the continuing activity of key members of the team.
- The research into diabetes conducted in 1981 by George Alberti and his team at Newcastle, and collaborators elsewhere, was selected to provide the case study for this preliminary stage for several reasons. It was thought to have been important science and there was a belief that some of it had made a contribution to clinical practice.

Chapter 2 : Bibliometric analysis

- An original plan to look at publications produced over a three year period was changed to looking at the output of just one year, 1981, because in that year alone Alberti and colleagues published 29 articles. These form the 1st generation papers and the average number of citations they received is high. Identifying the citations given to these 29 papers resulted in 799 2nd generation papers and 12,891 3rd generation papers. The numbers involved meant that it was impractical to go beyond the 3rd generation. Within the high overall average, the variation in the number of citations per paper was

considerable going from 76 to just one. Similarly, the half-lives of the 29 papers, ie the time taken for an article to receive 50% of its citations, ranged from two years to 11.

- Articles can be given a Research Level (ie one of four levels from clinical observation to basic) based on the journals in which they appear. Such analysis demonstrates the breadth of Alberti's work because the 29 articles are spread across all four Research Levels. Crucially, there was not a shift from basic to more clinical levels across the generations. The higher than average number of authors and addresses per paper is testimony to Alberti's extensive collaborations.
- The funding acknowledgements reveal the high proportion of papers supported, at least partially, by one funder: the British Diabetic Association, now Diabetes UK, which provided core support for Alberti's Newcastle team.

Chapter 3 : Categorisation of citations

- Traditional citation analysis does not allow identification of the importance of the cited article to the citing article, and therefore limits the ability to use citation analysis to trace the impact of basic or early research on later research. We conducted a review of the literature of the meaning of citations.
- From this review, a template was devised that allowed the location, nature and importance of citations to be recorded as well as the type of research (basic or clinical) described in the paper. This was used by six assessors on a sample of papers and inter-rater reliability was tested. Further work is required to refine the template and its definitions, and to improve its consistency in application.
- Nevertheless, for initial analysis, it was applied to 623 out of the 799 2nd generation papers. A four point scale was used for the importance of the cited paper to the citing paper. In just 9% of cases was the cited 1st generation paper thought to be in one of the top two categories, ie of Considerable or Essential importance to the citing paper.
- Statistical analysis revealed no relationship between the number of citations a paper received and the proportion of citations where the cited paper was classified as being of high (ie. Considerable or Essential) importance to the citing paper. Self-citations, however, were shown to be significantly more likely to be in this category.
- The classification of the type of research (basic or clinical) by our analysis of each paper broadly agreed with the classification of the journals by Research Level.
- The time constraints involved in applying the template, plus the lack of any overall pattern in terms of correlations between number and importance of citations, might point to the desirability of adopting a more selective approach, guided by qualitative analysis. In any selective approach, however, it is likely that self-citations should feature.

Chapter 4 : Qualitative analysis

- Given the number of co-authors, it seemed appropriate to send them a questionnaire rather than attempt to interview them. Therefore the interviewing was rather more concentrated than originally intended. Only one formal critical pathway was created, but it was undertaken by an expert in the field who worked with Alberti at Newcastle.
- Some problems emerged in taking 1981 as the starting point for the study. Alberti identified 10 selected papers from the 1970s and 1980s that he felt had had most impact on clinical practice. These helped to give us both a better understanding of the payback from our 1st generation, or 1981, papers, and provided further material for analysis.
- Attempting to describe the impact from the 1981 body of work, and from the 10 selected papers, underlines the complex reality of how science advances and influences clinical practice. If they make a contribution at all, most studies make a small, incremental one.
- A few papers, however, have been shown to have a considerably greater impact. A possible key to the level of payback indicated is the enormous breadth of Alberti's contacts, and fields and methods of working, to which various references were made. This is well illustrated in the account of how the idea for subcutaneous pumps came about. Similarly, the ability to produce the very important guidelines on treating diabetics during surgery, and diabetic coma, partly resulted from the application to clinical problems of the understandings gained from some of the basic/early clinical studies. It is significant that the key papers on these issues, all of which come from the list of 10 selected papers from the 1970s and 1980s, were having an impact on the 1981 work.
- How far the collection of papers from 1981 have been drawn upon in similar ways is less clear. Nevertheless, papers on treating diabetics during open heart surgery, and on bolus delivery of insulin at meal times, were key parts of these wider streams, despite variable citation levels. Furthermore, various papers, including on acarbose, on portal infusion of insulin, and on semi-human insulin, were important steps in bodies of work in their respective areas. The complexity was illustrated by a paper that helped debunk the Chlorpropamide alcohol flushing hypothesis, and thus end a line of scientific enquiry: there was payback in stopping an incorrect line of inquiry, but nothing on which to build.
- Each technique in the qualitative study produced information about the successful subsequent careers followed by many researchers trained through working with Alberti.
- Historical perspectives, and insider expert opinions, were important in the qualitative analysis. Overall, the qualitative methods highlighted some limitations in the bibliometric approach but also showed how aspects of the citation analysis can complement the opinions expressed, for example about the importance of the breadth of Alberti's work.

Chapter 5 : Lessons learnt and the way forward

- **Lessons learnt:** a variety of methods can be used successfully to gather considerable data about the payback from a body of research undertaken 20 years ago. Traditional citation analysis alone, however, is not sufficient: the importance of the surgery papers despite their relatively low citation rates illustrates this. The qualitative methods are important and much of the analysis is strengthened by drawing on multiple approaches. Several problems remain, including: identifying a coherent starting point for the analysis; coping with the enormous number of papers involved in later generations; and refining the template for categorising citations and developing ways of fully utilising the results from applying it.
- **Preparing for the large-scale study:** this preliminary study provides a basis on which to attempt to undertake the larger study we envisaged. Issues now being addressed include identification of the level of bibliometric/citation analysis necessary to complement any qualitative studies. To provide confidence in the findings from an eventual large-scale study, we will need to expand the focus. The study will need to cover at least four sets of case studies. Ideally, each set should focus on a number of research groups working in a country in the same field. We hope there will be sets of case studies in two or three fields and in at least two countries. The issues to be explored will include ones highlighted by this study such as breadth of work, level of collaboration, and the role of core funding.
- **Methods for the large-scale study:** for each case study we now propose to employ two methodological elements based on the qualitative and quantitative techniques adopted in the preliminary study. They will work in parallel but the quantitative bibliometric analysis would be applied selectively to parts of 'research lines' (ie discrete themes of research) identified in the qualitative studies as being important in influencing clinical practice.
- **Presenting the findings:** each research line could be written-up in a standardised document that would use the HERG payback model and categories to describe the impact of that research. We shall use the qualitative and quantitative data to compare and contrast the 'payback' of research lines by country and disease, and then identify common factors that correlate with the translation of basic or early clinical research.
- **Concluding comments:** in the era of 'evidence based policy', research funders are looking for value for money in the research they support and for evidence on the effectiveness of different research strategies. In this study we have begun developing a methodology that will allow us to understand the complexity of research development over a series of generations. The utility of the policy research we propose here will only be realised when it is scaled up to cover a number of different fields in different settings.

1 INTRODUCTION

1.1 Background

How to assess the utilisation or impact of health research, and the payback or benefits from it, is an issue of growing importance to national and international bodies (Raiten and Berman, 1993; Buxton and Hanney, 1998; National Institutes of Health, 2000; COHRED, 2000; Smith, 2001; WHO, 2002; van Weel, 2002; Royal Netherlands Academy of Arts and Sciences, 2002; Pang *et al*, forthcoming). A number of studies have now shown that it is possible to make an *ad hoc* assessment of the payback, or impact on policy or practice, of specific examples of applied 'health services research' (Buxton and Hanney, 1994, 1996 and 1997; Buxton *et al*, 1999; Jacob and McGregor, 1997). Increasingly, the view is being adopted that such approaches should form part of the regular process of assessing research and thus contribute to the justification for further funding (Buxton, Croxson and Hanney, 1999; Croxson *et al*, 2000; Smith, 2001).

There remains, however, a belief that these techniques cannot be applied or adapted to address the much more long-term and indirect impact that basic science or early clinical research may eventually have on health service practice and policies (Cozzens, 1997). This in turn leads to a dilemma. One option is that basic research will be exempt from consideration of its impact, so making the 'quality of its science' the only criterion for such work. Such a position would negate much of the logic of such initiatives as the UK Foresight Exercise (OST, 1993) and may give, or appear to give, an unfair advantage to basic research relative to more applied research. Alternatively, the impact of basic science will be measured in some way but without any real understanding of how impact occurs, or how it might be best assessed: such an uninformed situation could potentially disadvantage more basic science research relative to applied research, or lead to a distortion of priorities.

There has been the occasional one-off study, including an attempt to integrate cost-benefit analysis with historical tracing of the important science developments leading up to the discovery of the methodology for producing monoclonal antibodies (Raiten and Berman, 1993). Furthermore, the National Institutes of Health (NIH), as part of their report under the Government Performance and Review Act, take various approaches to accounting for the way in which their funding has been spent. They not only provide brief accounts of specific scientific findings described in individual papers linked to NIH funding, but also two page vignettes describing the longer history of certain issues and recent advances (NIH, 2000). These are called 'Stories of Discovery'. These stories do not amount, however, to a systematic attempt to assess the payback from research.

It has been suggested that clinical guidelines can be viewed as one form of secondary output from health research (Buxton and Hanney 1996). Taking this concept, it is possible to study the nature of the evidence base of such guidelines by examining the citations included in them, and then, in turn, to consider the citations in those papers (Grant, 1999; Grant *et al*, 2000 - see also Lewison and Wilcox-Jay, 2003). This form of working backwards from desirable impacts of the research partially builds on the classic study conducted in the USA in the 1970s. That study entailed working backwards from then recent clinical advances to uncover the research that lay behind them (Comroe and Dripps, 1974; Comroe and Dripps, 1976). Although a one-off and imperfect study, 'Comroe and Dripps' is often cited as an irrefutable indication of the breadth and unpredictability of how basic science will eventually be applied. Attempts to replicate the Comroe and Dripps study (Mason *et al*, 2001; Grant *et al*, 2003) suggest that starting with clinical advances, and attempting to understand the linkages back to basic research, involves too many assumptions.¹

It has been proposed instead that our understanding of this important area would be valuably informed by undertaking studies that start with the more basic, or early clinical, science and try to trace them through to applications (Buxton and Schneider, 1999). Along similar lines, a major review of the role of citations suggested that:

'One largely unutilized role of citations is to serve as a "radioactive tracer" of research impacts. Citations allow the analyst to track the documented flow and evolution of research over time until the linkages to far downstream products can be identified. Because of the potential information available from the tracking application, this is a very fruitful area for future citation research and analysis'. (Kostoff, 1998, pp 29-30).

For this current project, we identified various approaches for undertaking this process of looking forward (with hindsight). These included an essentially quantitative analysis of literature and citations involving both mechanical tracing of citations and a categorisation of the citations found, and a more qualitative approach drawing on the opinions of researchers and experts through questionnaires and interviews.²

¹ Diabetes UK also explored with the Policy Unit of the Wellcome Trust the possibility of undertaking a study that worked backwards from current clinical advances. They, too, found that it soon became apparent that the number of papers involved would be overwhelming.

² Support for the integration of scientometric methods into a qualitative study of science has come recently from Gläser and Laudel (2001), but they examined a rather different issue than the one of concern here.

1.2 Potential large-scale study

Our medium-term objective is to undertake a large study that would use these approaches to trace the impact of a number of discrete bodies of clinical/basic research, probably identified in terms of a body of funded research undertaken by an individual or clearly identifiable research group. The aim would be to assess the impact of several such bodies of research, probably work undertaken about twenty years ago funded by research agencies in the UK and abroad. We envisage that a number of funding agencies whose funded research would be studied would be approached jointly to fund such a large-scale study.

This research could make a major contribution to our understanding of the way that results of basic or clinical research (specific findings, ideas, concepts or methodological approaches) eventually feed through into application. It might thus provide a basis for proposing indicators of likely future impact, say in terms of the nature of the citations received and the type of journals in which they occur.

1.3 Preliminary study

Before embarking on any such large-scale multi-agency study, it was important to test the key components of our proposed methodology. It was thought that if the various elements were undertaken together they would be mutually supportive, help to establish the feasibility of a major study, and identify the most appropriate methods, scope, costs and time-scale. Additionally, it was hoped that if this preliminary study focussed on a body of research of real interest and import, it could also provide an interesting case study in its own right.

The preliminary study reported here was funded by the London Regional Office of the NHS whose Director of R&D, Sally Davies, suggested diabetes or cardiology might be appropriate areas in which to conduct the study. It was judged that we needed to follow research for at least 20 years. Initial reviews quickly established two bodies of research in diabetes that had been conducted 20 years ago that might be suitable. Of the two teams, the one based at Newcastle and led by George (now Sir George) Alberti was identified as most suitable because of the importance of the work and the belief that some contribution to clinical practice had occurred. The intention was that papers from the three years 1981-83 would be studied, but the high productivity of Alberti and his various collaborators, and the high average level of citations to the papers, meant that it rapidly became clear that the preliminary study could focus on only one year. The base year, 1981, was chosen.

The study was conceived as having three main elements and these are considered in turn in each of the next three chapters:

- standard bibliometric analysis;
- categorisation of citations;
- qualitative analysis.

The scale of the bibliometric analysis described in Chapter Two was immense. The various techniques were not just applied to the initial body of 29 papers from 1981 (the 1st generation) but also to papers that cited them. The citations were traced through several generations and numbered almost 13,000 in the 3rd generation.

Citation analysis is widely undertaken but nevertheless subject to frequent criticism. Part of the unease is the lack of certainty about what meaning can be attached to citations and thus to citation scores. Chapter Three starts by describing a variety of previous studies that have attempted to analyse citation behaviour with the aim of better understanding its meaning. The review was used to inform the development of a template to be applied to the citations. Whilst this was not as comprehensive as some previous categorisations, it included the features central to our focus on the importance of the cited paper to the citing author. The application of this template resulted in the range of information described in Chapter Three.

In parallel to the first two elements, we wished to trial a more formal qualitative approach to understand the processes by which the body of research influenced subsequent work, indirectly facilitated applied research, and perhaps contributed to an applied impact. It was hoped this would add qualitative understanding to the first two elements of the work. Chapter Four describes this qualitative approach. A combination of critical pathway analysis by experts in the research area and qualitative methods of questionnaires and interviews were used. The interviewing and the subsequent analysis adopted the triangulation techniques of building on the data from various methods and sources and comparing accounts.

It was recognised from the outset that it would be extremely difficult to identify the impact of one set of papers from a short time period. Furthermore, the preliminary study would be more complex and resource intensive than any eventual large-scale study because a range of methodological tools had to be developed and tested. As it turned out, the study was even more complex, but also potentially more instructive, than originally envisaged as issues were addressed that had not been in the original plan. Some of these arose from the advice of the project's Advisory Group whose membership is listed in the acknowledgements. Others resulted from the extremely rich and varied nature of the body of work examined.

Alberti, more than many researchers, believes in interaction between basic and clinical research. As he told us when interviewed: 'you raise questions in humans you can't answer; you go to the animals to try to answer them; you go back to the humans to confirm and develop them...yin and yang'. This approach, together with the many and diverse collaborations in which Alberti was involved, meant that the study was less able than anticipated to focus on following one clear strand of work, but made more progress in identifying clinical impact. Such impact, however, could not neatly be associated with one year's work. Alberti, therefore, made various helpful suggestions which whilst they could not be substantially followed up in the first two elements were partially brought into the qualitative analysis.

The final chapter builds on the previous ones to draw out the implications as to how further work could be conducted in the large-scale study, and to make proposals for that study.

2 BIBLIOMETRIC ANALYSIS

2.1 Introduction

This chapter presents the results of a bibliometric analysis, using the Science Citation Index (SCI) of the Institute for Scientific Information (ISI), of the data gathered about the three generations of papers starting with Alberti's output from 1981 as the 1st generation. Information relating to all three generations mapped in the study have been analysed and compared. Factors covered include Research Levels, numbers of authors, their addresses, and numbers of addresses. For the 1st generation publications the analysis also includes half-lives and funding sources. The methods used are standardised and the results will be compared with previous studies in the field of biomedicine.

2.2 Methods

2.2.1 Citation tracing and mapping

The output of Alberti from 1981 was obtained from the SCI. The 1981 SCI CD-ROM version was searched for publications from ALBERTI K* with Newcastle in the address field. This constituted the 1st generation output and consisted of 49 publications in total; 22 meetings abstracts, 1 letter and 26 articles. Standard bibliometric analyses use only papers classified by the SCI as either articles, notes or reviews. On consultation with Philip Home, we realised that three papers were missing from the SCI-generated list, all from the journal *Diabetes Care*. The details of these articles were available from MEDLINE and therefore added to our database of 1st generation publications. The 1st generation output was stored on an Excel worksheet, with each paper given a unique identifying code (Table 1). Papers A1-A26 are recorded in alphabetical order by journal, and papers A27-A29 represent the additional *Diabetes Care* papers added subsequent to the SCI search. When the Research Outputs Database (ROD) filter DIABE, which is designed to identify papers relating to diabetes, was applied to the 1st generation papers, eight of them (A2, A3, A9, A10, A11, A19, A20, A25) were assessed as not being diabetes-related. However, on further examination, four (A2, A3, A10 and A25) were clearly diabetes-related and at least one of them (A3) was viewed by one of the co-authors as being important in later research. This highlights the difficulty of using a bibliometric tool such as a subject filter in a small-scale study. We therefore took the whole 29 papers from Alberti in 1981 as the set on which to base analysis and comparisons.

Table 1: 1st generation papers

Ref No	Authors	Title	Journal	Year	Volume	Pages
A1	Madsbad-S Faber-OK Binder-C Alberti-KGMM Lloyd-B	Diurnal Profiles of Intermediary Metabolites in Insulin-Dependent Diabetes and Their Relationship to Different Degrees of Residual B-Cell Function	ACTA DIABETOLOGICA LATINA	1981	Vol 18	pp 115-121
A2	Burrin-JM Worth-R Law-S Alberti-KGMM	Simple Filter-Paper Method for Home Monitoring of Blood-Glucose, Lactate, and 3-Hydroxybutyrate	ANNALS OF CLINICAL BIOCHEMISTRY	1981	Vol 18	pp 243-247
A3	Conlon-JM Whittaker-J Hammond-V Alberti-KGMM	Metabolism of Somatostatin and Its Analogs by the Liver	BIOCHIMICA ET BIOPHYSICA ACTA	1981	Vol 677	pp 234-242
A4	Whittaker-J Hammond-VA Alberti-KGMM	Effects of Colchicine on Insulin Binding to Isolated Rat Hepatocytes	BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS	1981	Vol 103	pp 1100-1106
A5	Gill-GV Sherif-IH Alberti-KGMM	Management of Diabetes During Open-Heart Surgery	BRITISH JOURNAL OF SURGERY	1981	Vol 68	pp 171-172
A6	Barnett-AH Spiliopoulos-AJ Pyke-DA Stubbs-WA Burrin-J Alberti-KGMM	Metabolic Studies in Unaffected Co-Twins of Non-Insulin-Dependent Diabetics	BRITISH MEDICAL JOURNAL	1981	Vol 282	pp 1656-1658
A7	Owens-DR Jones-MK Hayes-TM Heding-LG Alberti-KGMM Home-PD Burrin-JM Newcombe-RG	Human Insulin - Study of Safety and Efficacy in Man	BRITISH MEDICAL JOURNAL	1981	Vol 282	pp 1264-1266
A8	Pickup-JC Home-PD Bilous-RW Keen-H Alberti-KGMM	Management of Severely Brittle Diabetes by Continuous Subcutaneous and Intramuscular Insulin Infusions - Evidence for a Defect in Subcutaneous Insulin Absorption	BRITISH MEDICAL JOURNAL	1981	Vol 282	pp 347-350
A9	Mcculloch-AJ Johnston-DG Burrin-JM Hodson-AW Clark-F Waugh-C Orskov-H Alberti-KGMM	Diurnal Hormone-Metabolite Profiles in Hypothyroidism	CLINICAL ENDOCRINOLOGY	1981	Vol 15	pp 607-619
A10	Blesa-Malpica-G Johnston-DG Burrin-JM Orskov-H Heath-AB Alberti-KGMM	Dopaminergic Control of Ketogenesis in Fasting	CLINICAL ENDOCRINOLOGY	1981	Vol 14	pp 479-484
A11	Harvey-JE Baldwin-CJ Wood-PJ Alberti-KGMM Tattersfield-AE	Airway and Metabolic Responsiveness to Intravenous Salbutamol in Asthma - Effect of Regular Inhaled Salbutamol	CLINICAL SCIENCE	1981	Vol 60	pp 579-585

Ref No	Authors	TitleWhen the	Journal	Year	Volume	Pages
A12	Massi-Benedetti-M Noy-G Johnston-IDA Worth-R Alberti-KGMM	Glucose Controlled Insulin Infusion System (Biostator) Application During Surgery for a Presumed Pancreatic Micro-Insulinoma	DIABETE & METABOLISME	1981	Vol 7	pp 41-44
A13	Jenkins-DJA Taylor-RH Goff-DV Fielden-H Misiewicz-JJ Sarson-DL Bloom-SR Alberti-KGMM	Scope and Specificity of Acarbose in Slowing Carbohydrate-Absorption in Man	DIABETES	1981	Vol 30	pp 951-954
A14	Nosadini-R Noy-G Kurtz-AB Alberti-KGMM	Differential Response to Infusions of Highly Purified and Conventional Bovine and Porcine Insulins	DIABETES	1981	Vol 30	pp 650-655
A15	Whittaker-J Cuthbert-C Hammond-V Alberti-KGMM	Impaired Insulin Binding to Isolated Adipocytes in Experimental Diabetic-Ketoacidosis	DIABETOLOGIA	1981	Vol 21	pp 563-568
A16	Andrews-WJ Henry-RW Alberti-KGMM Buchanan-KD	The Gastroenteropancreatic Hormone Response to Fasting in Obesity	DIABETOLOGIA	1981	Vol 21	pp 440-445
A17	Nosadini-R Noy-GA Alberti-KGMM Hodson-A Orskov-H	The Metabolic Response to Hyperglycemic Clamping in Insulin-Dependent Diabetes	DIABETOLOGIA	1981	Vol 20	pp 113-117
A18	Metcalfe-P Johnston-DG Nosadini-R Orskov-H Alberti-KGMM	Metabolic Effects of Acute and Prolonged Growth-Hormone Excess in Normal and Insulin-Deficient Man	DIABETOLOGIA	1981	Vol 20	pp 123-128
A19	Hanson-RD Gray-RM Alberti-KGMM	Liver Metabolites in Resting and Exercising Rats at 1-Bar and 4-Bar	JOURNAL OF APPLIED PHYSIOLOGY	1981	Vol 51	pp 1326-1330
A20	Twentyman-OP Disley-A Gribbin-HR Alberti-KGMM Tattersfield-AE	Effect of Beta-Adrenergic-Blockade on Respiratory and Metabolic Responses to Exercise	JOURNAL OF APPLIED PHYSIOLOGY	1981	Vol 51	pp 788-793
A21	Schmitz-O Alberti-KGMM Hreidarsson-AB Laurberg-P Weeke-J Orskov-H	Suppression of the Night Increase in Serum TSH During Development of Ketosis in Diabetic-Patients	JOURNAL OF ENDOCRINOLOGICAL INVESTIGATION	1981	Vol 4	pp 403-407
A22	Owens-DR Jones-MK Hayes-TM Heding-LG Alberti-KGMM Home-PD Burrin-JM	Comparative-Study of Subcutaneous, Intramuscular, and Intravenous Administration of Human Insulin	LANCET	1981	Vol 2	pp 118-122
A23	Desilva-NE Tunbridge-WMG Alberti-KGMM	Low Incidence of Chlorpropamide-Alcohol Flushing in Diet-Treated, Non-Insulin-Dependent Diabetes	LANCET	1981	Vol 1	pp 128-131
A24	Stevenson-RW Parsons-JA Alberti-KGMM	Comparison of the Metabolic Responses to Portal and Peripheral Infusions of Insulin in Diabetic Dogs	METABOLISM-CLINICAL AND EXPERIMENTAL	1981	Vol 30	pp 745-752

Ref No	Authors	Title	Journal	Year	Volume	Pages
A25	Nosadini-R Alberti-KGMM Johnston-DG Del Prato-S Marescotti-C Duner-E	The Anti-Ketogenic Effect of Alanine in Normal Man - Evidence for an Alanine-Ketone Body Cycle	METABOLISM-CLINICAL AND EXPERIMENTAL	1981	Vol 30	pp 563-567
A26	Home-PD Pickup-JC Keen-H Alberti-KGMM Parsons-JA Binder-C	Continuous Subcutaneous Insulin Infusion - Comparison of Plasma-Insulin Profiles After Infusion or Bolus Injection of the Mealtime Dose	METABOLISM-CLINICAL AND EXPERIMENTAL	1981	Vol 30	pp 439-442
A27	Massi-Benedetti-M Burrin-JM Capaldo-B Alberti-KGMM	A Comparative Study of the Activity of Biosynthetic Human Insulin and Pork Insulin Using the Glucose Clamp Technique in Normal Subjects	DIABETES CARE	1981	Vol 4	pp 163-167
A28	Taylor-R Home-PD Alberti-KGMM	Plasma Free Insulin Profiles After Administration of Insulin by Jet and Conventional Syringe Injection	DIABETES CARE	1981	Vol 4	pp 377-379
A29	Worth-R Harrison-K Anderson-J Johnston-DG Alberti-KGMM	A Comparative Study of Blood Glucose Test Strips	DIABETES CARE	1981	Vol 4	pp 407-411

The SCI was searched from 1981 to June 2001 for the papers that cited the 29 1st generation papers, and thus constituted our 2nd generation output. The format for the search was as follows³:

1st author-year*-vol-1st page number

eg FRAME-IA-1983*-V34-P51

* is a wild-card entry to catch all journals recorded with the SCI.

In addition to identifying the numbers of citations in each generation, the relationships between papers in the generations were also mapped. In all, there were 848 2nd generation publications of which 799 were classified as articles notes and reviews. (The remainder were items such as letters). The 2nd generation articles, notes and reviews were given an identifying number and linked directly to the 1st generation paper(s) that they cited. As a result of some 2nd generation publications citing more than one 1st generation paper, the number of unique 2nd generation papers was lower at some 730. The details of these unique 2nd generation publications were used to search the SCI CD-ROM database (1981–June 2001) for 3rd generation papers ie. those papers citing the 2nd generation papers. This resulted in a total of 13,542 3rd generation papers of which 12,891 were classified as articles, notes or reviews (comprising 9,376 unique papers). Four hundred and seventy-four of these 3rd generation papers were also 2nd generation publications.

2.2.2 Research Levels

Research Level values for many journals have been determined by CHI Research Inc (Narin *et al*, 1976). The system is based on expert opinion and journal-to-journal citations. Journals are allocated into four hierarchical levels in which each level is more likely to cite papers in journals at the same level or the level below and *vice versa* (Table 2).

³ This format failed to give any information about paper A29 which appeared to have no citations and the format was altered to include *Diabetes Care* in place of the *. This yielded 28 papers, of which 24 were articles, notes or reviews. This is an important point to note regarding the search strategy and may have led to some citations of other papers being missed.

Table 2: Definition of research levels

Research Level	Type	Example
1	Clinical observation	<i>British Medical Journal</i>
2	Clinical mix	<i>New England Journal of Medicine</i>
3	Clinical investigation	<i>Immunology</i>
4	Basic research	<i>Nature</i>
N/A or 0	Yet to be classified/difficult to classify	-

According to the CHI Research definitions only 4% of papers in Research Level (RL) 1, ie clinical observation journals (eg *British Medical Journal*) will cite papers in RL4, basic journals (eg *Nature*), compared to 8% for RL2 clinical mix journals, (eg *New England Journal of Medicine*), and 21% for RL3 clinical investigation journals (eg *Immunology*). (Wellcome Trust/NHS Executive, 2001).

The Research Level, in theory, provides us with a transparent, systematic and generally accepted method of identifying basic research. However, there are two caveats. This method of categorisation assumes that: 1) Research Levels have remained constant over time (20 years in the case of this study); and 2) all papers published within a journal are of the same Research Level.

2.3 Results

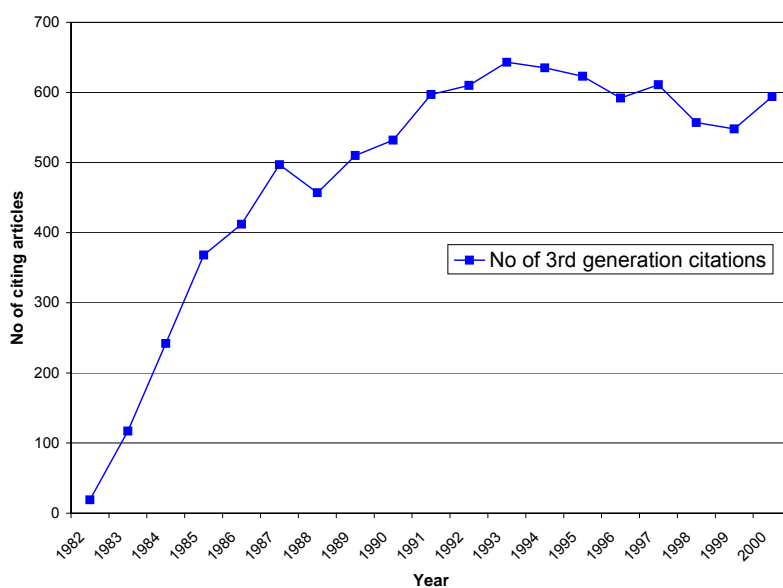
2.3.1 Number of papers and self-citations

The total number of citations per year in the 2nd generation is shown in Figure 1. The peak of the 2nd generation papers occurs in 1983, two years after the publication of the original articles and then falls gradually until 2001. The number of citations in the 3rd generation, per year, is shown in Figure 2. The number of papers rises steeply from 1983 through to 1987 where there is a short dip. This is followed by a second rise through to 1993. Interestingly, there is a ten-year gap between the peaks in each generation.

Figure 1: Number of 2nd generation papers per year

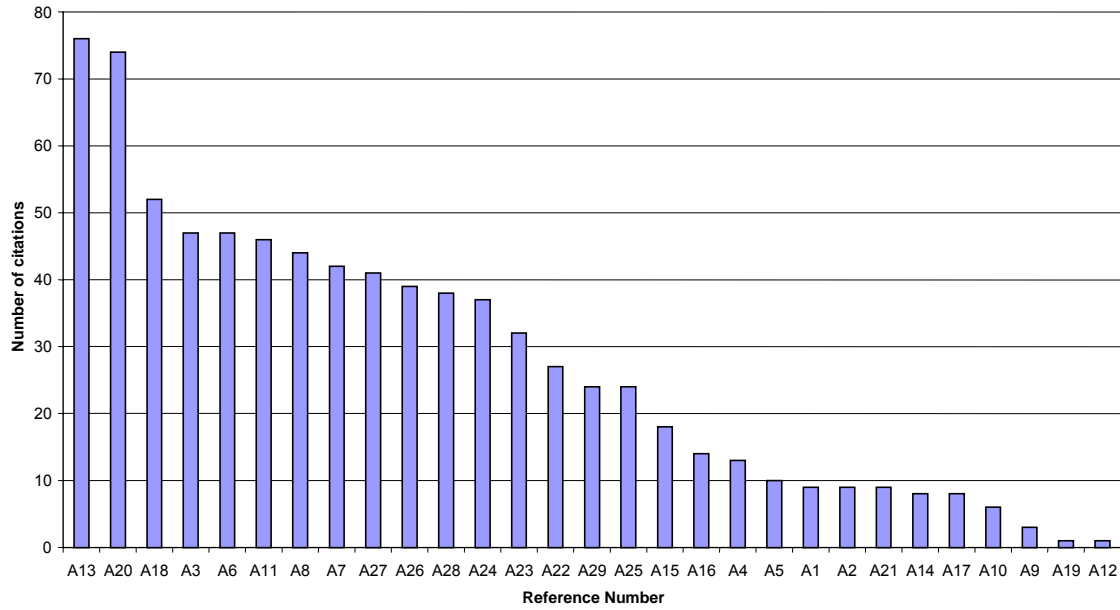


Figure 2: Number of 3rd generation papers per year



The number of citations for each of the 1st generation papers (articles, notes and reviews only) is shown in Figure 3. Two papers (A13 and A20) were cited on more than 70 occasions (76 and 74 times respectively). Sixteen papers were cited between 10 and 52 times with the remaining papers cited less than 10 times during the 20-year period. There were 623 2nd generation papers for which we have recorded information. Of those papers 114 were self-citations where one or more of the authors of the 2nd generation publication was named on the cited paper. The number of citations given to each of the 2nd generation papers ranged from none to 469.

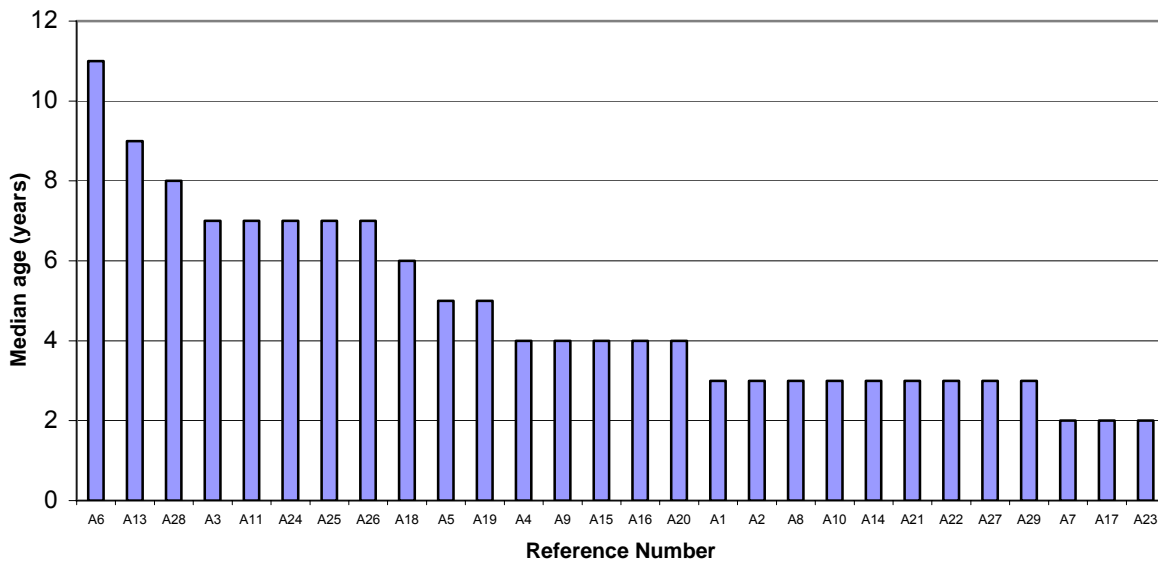
Figure 3: Citations of 1st generation papers



2.3.2 Half-lives

Generally, a research article is born, matures in use, produces new information, loses value, becomes obsolete and eventually dies (Pemberton and Nugent, 1995). The measure of a journal article’s half-life is therefore a rather crude measure of its value and obsolescence. Half-life is measured by the time it takes for an article to receive 50% of its citations. The half-lives of the 29 1st generation papers are shown in Figure 4.

Figure 4: Half-lives of 1st generation papers

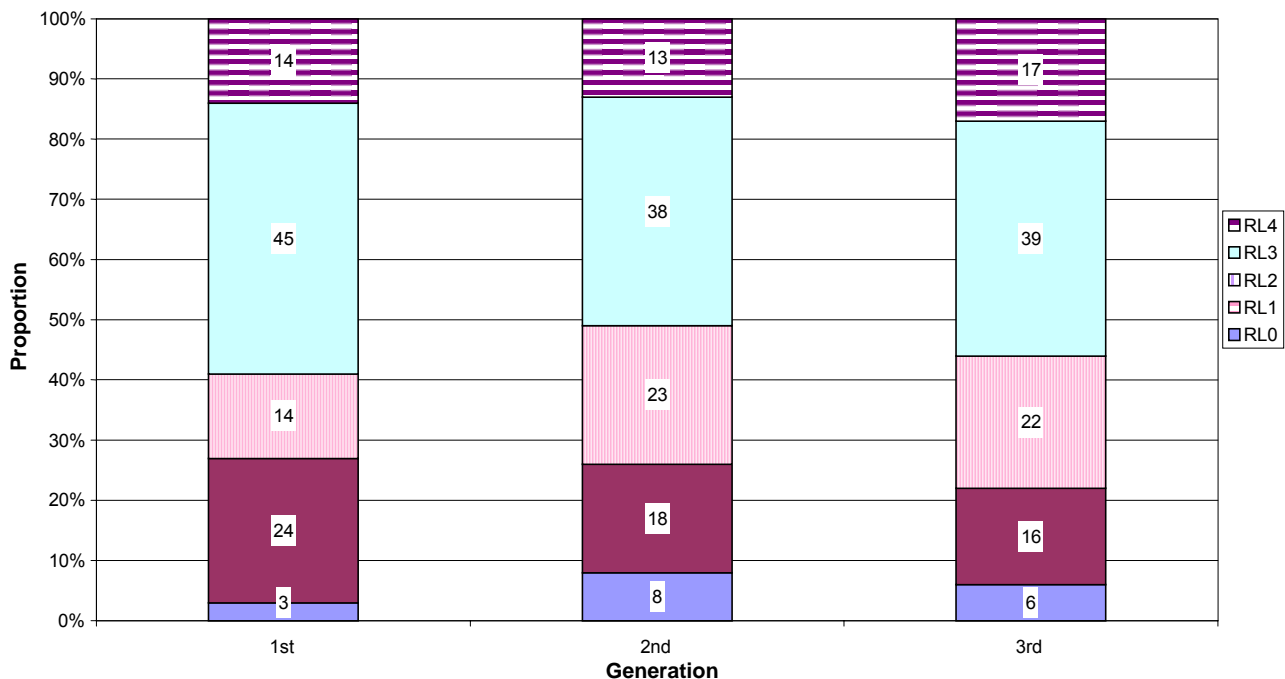


The half-life for the 29 1st generation articles taken as a group was five years and the half-life for the 2nd to 3rd generation output was seven years.

2.3.3 Research Levels

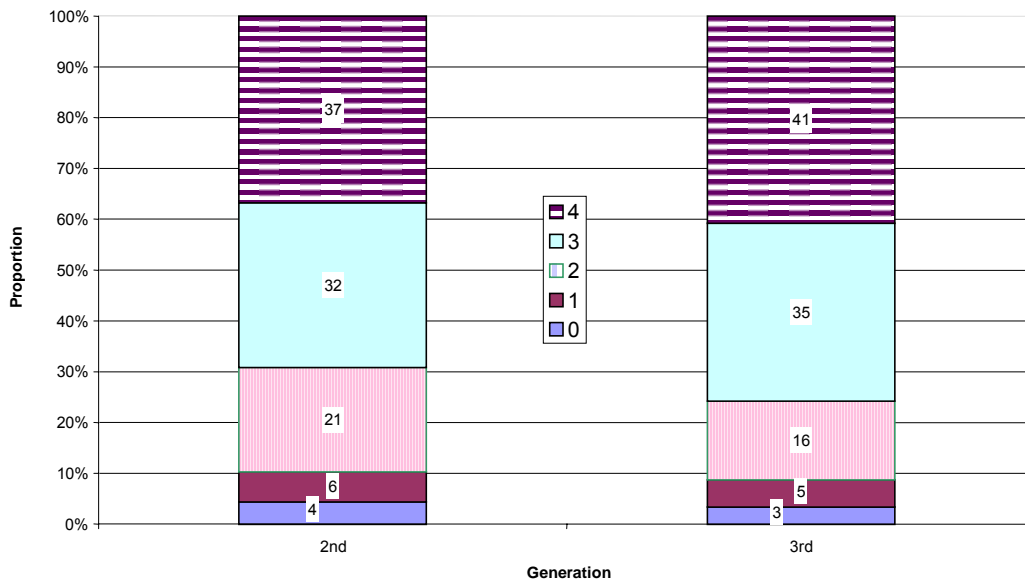
The Research Level (RL) of the papers in each of the three generations is shown in Figure 5. The RLs, as noted on Table 2, go from clinical observation (RL1) to basic research (RL4), but no Research Level had been allocated to a small number of the journals. There is a broad similarity in the proportion of papers within each Research Level between generations, with no apparent shift, as might have been expected, from RL4 to RL1. These papers are slightly more clinically oriented than the proportions published previously for all diabetes research in England: 13% RL1, 18% RL2, 42% RL3 and 22% RL4. (Wellcome Trust/NHS, 2001).

Figure 5: Research Levels of 1st, 2nd and 3rd generation papers



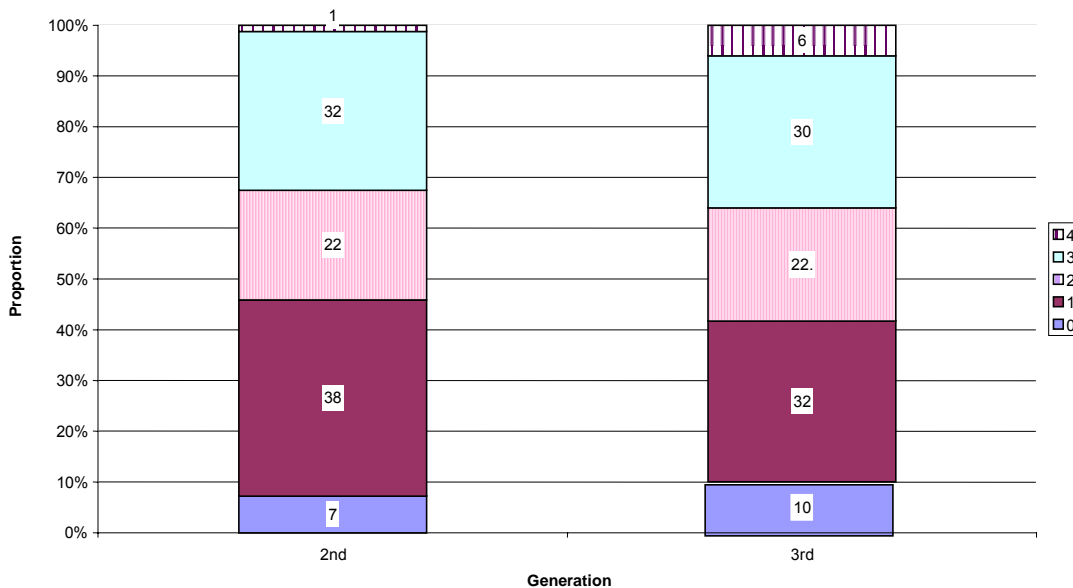
When only RL4 papers (basic research) from the 1st generation were analysed, the distribution of Research Levels of the papers citing them in the 2nd and 3rd generations was different from the overall picture: there were proportionately more RL4 papers and fewer RL1 papers (Figure 6).

Figure 6: 2nd & 3rd generation RLs from RL4 1st generation papers



Similarly, when only RL1 papers from the 1st generation were used to define subsequent generation outputs, the papers remained within the clinical (RL1 and RL2) journals (Figure 7). This finding supports the basis of the CHI Inc Research Level system where papers in journals of a given Research Level are more likely to cite papers in journals at the same level or the level below and *vice versa*.

Figure 7: 2nd & 3rd generation RLs from RL1 1st generation papers



2.3.4 Number of authors and addresses

The proportion of papers with different numbers of authors between the three generations is shown in Figure 8. None of the 1st generation papers had less than 3 authors, with 17 of 29 of the papers having five or more authors, much higher than in subsequent generations. This appears to indicate that Alberti was collaborating with a large number of people and/or that he had a large active research group at that time. Recent figures show a national average in biomedical research of 3.8 authors per paper (The Wellcome Trust, 1998). They also show, however, that between 1988 and 1995 the percentage of papers published in the UK with over 5 authors gradually increased, indicating that there was an increasing level of collaboration between biomedical researchers. The average number of authors per paper in diabetes research in England between 1990 and 1997 was just over five (The Wellcome Trust/NHS, 2001).

Figure 8: Proportion of papers with different numbers of authors in all 3 generations

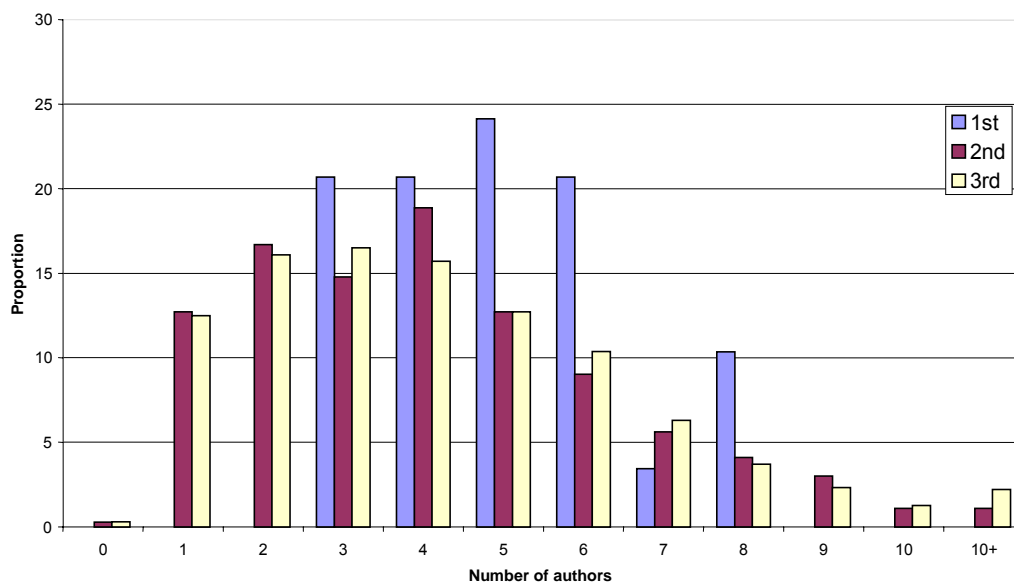


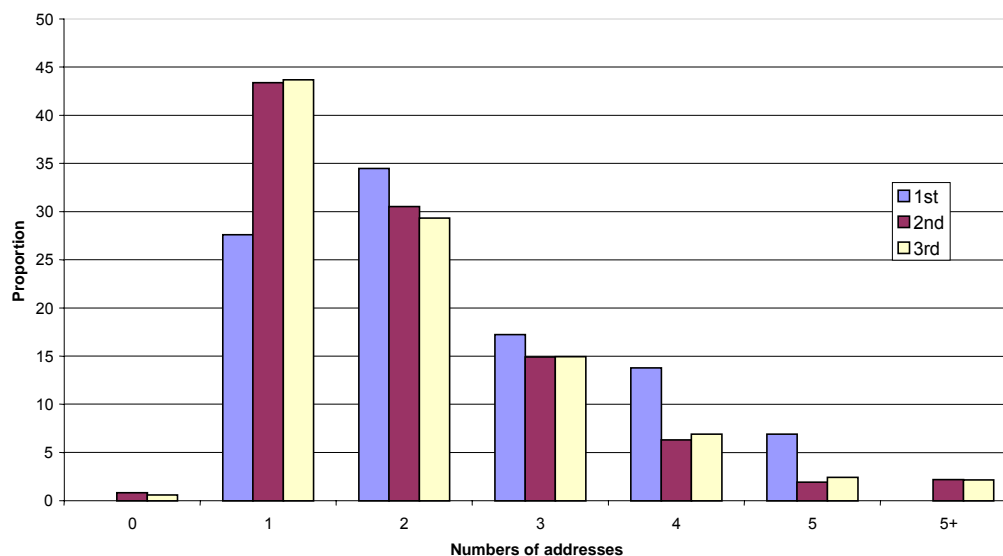
Table 3 summarises the median and mean numbers of authors for all papers in each of the three generations. The median number of authors is slightly higher for the 1st generation papers than either of the subsequent generations as is the mean.

Table 3: Mean and median number of authors for each of the three generations

	1 st generation	2 nd generation	3 rd generation
Mean ± sd	5.0 ± 1.6	4.1 ± 2.4	4.3 ± 5.3
Median	5	4	4

The number of addresses per paper is an indication of collaboration between different research groups. The numbers of addresses per paper for all three generations are shown in Figure 9.

Figure 9: Proportion of papers with different numbers of addresses in all 3 generations

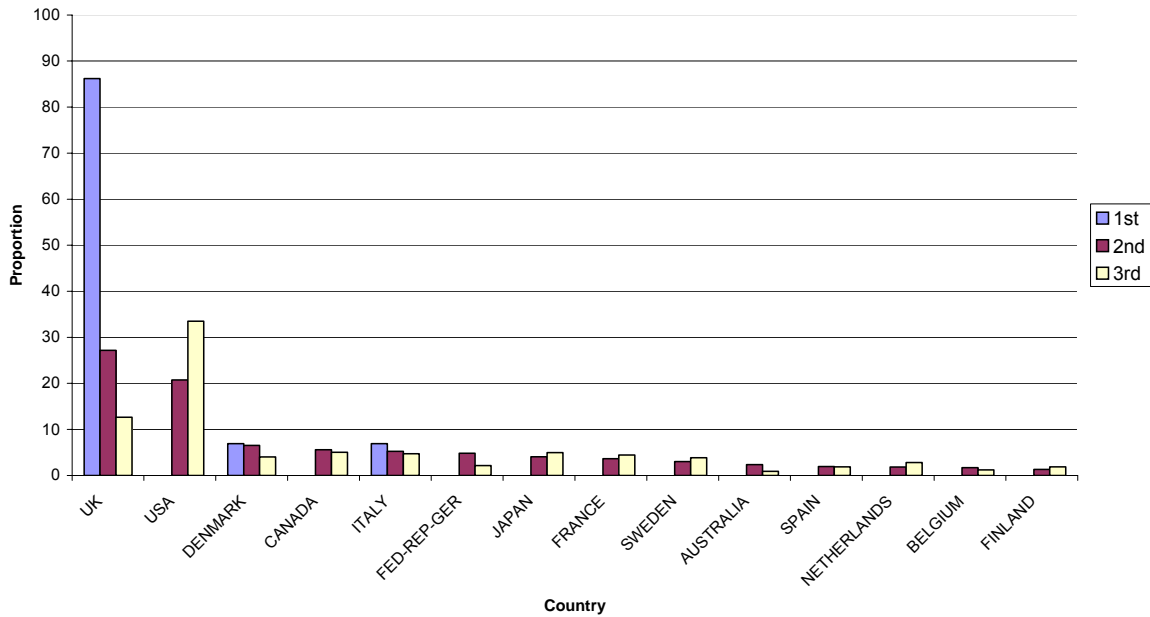


There are 28% of 1st generation papers with a single address, approximately half of those in subsequent generations (45% 2nd generation and 44% 3rd generation). The proportion of UK biomedical research papers in 1995 bearing single addresses was 44.3% (The Wellcome Trust, 1998). There are 14% of 1st generation papers with four addresses, which is, approximately, double those in subsequent generations (6% generation 2 and 7% generation 3). This suggests that Alberti's group was collaborating with other research groups at this time. The UK national average number of addresses per paper was 2.0 in 1995.

2.3.5 Authors' addresses by country and distribution in the UK

Figure 10 shows the proportion of papers in each of the three generations from different countries. The main point to note here is that 2nd generation publications continue to have a proportionately high level of UK addresses compared to the USA. However, by the 3rd generation, the number of UK addresses has fallen in comparison with the USA although it remains higher than the other countries examined.

Figure 10: Proportion of papers in each generation by country of 1st author



The distribution of all UK addresses for all three generations was examined. It indicates that in the first generation there was some residual output from Alberti's time spent at Southampton, in addition to ongoing collaborative research at Cardiff, Oxford and London. In the 2nd generation papers the proportion of Newcastle addresses remains high, presumably through the effect of self-citation, although of course a wider range of centres also feature. By the 3rd generation the influence has spread throughout the UK, but Newcastle, Oxford, London and Nottingham emerge as the centres with the highest output of papers containing citations to the 2nd generation papers.

2.3.6 Funding sources

The funding sources acknowledged in the 1st generation papers were examined and 20 of the 29 papers acknowledged financial support from the British Diabetic Association (Table 4), which provided core support for Alberti's group at that time. The Wellcome Trust and the UK Medical Research Council (MRC) were acknowledged on five and four papers respectively, with four of the 29 papers not known to have specific financial support. The actual financial input from each of those sources is not recorded here, but seventeen of the papers acknowledged two or more sources of support and details of the funding for each paper are given in Appendix 1.

Table 4: Funding sources acknowledged and/or identified for 1st generation papers

Funding Sources	Number of papers
British Diabetic Association	20
Novo Industries, Denmark	8
Wellcome Trust	5
Medical Research Council	4
British Council	2
British Insulin Manufacturers	2
Newcastle Area Health Authority	2
Newcastle Area Health Authority (Teaching)	2
Bayer UK	1
Danish Medical Research Council	1
Eli Lilly	1
Medistron Ltd and Boehringer Corporation Ltd [for the modified Glucocheck meter]	1
Minet Trust	1
National Institutes for Health	1
Northern Regional Health Authority	1
Pfizer Ltd [materials]	1
Royal Victoria Hospital Belfast Research Fellowship	1
Servier Laboratories	1
Synthelabo (Paris)	1
None	4

2.4 Key findings

2.4.1 Key observations about the body of work

The average number of citations to the 1st generation papers was higher than anticipated in our project application. This probably reflects the importance of Alberti's portfolio of work plus the fact that diabetes journals have one of the highest average impact factor scores (Wellcome Trust/NHS Executive, 2001). Within the high overall average, there was considerable variation in the number of citations. Similarly, the half-lives showed a wide range.

The Research Level of papers demonstrated the breadth of Alberti's work -- the 29 articles were spread across all four Research Levels. Crucially, there was not a shift from basic to more clinical levels across the generations. The information about the higher than average number of authors and addresses per paper is again testimony to the extent of Alberti's collaborations. The analysis of the nationality of the addresses shows, firstly, that Alberti was collaborating internationally and, secondly, how far the work spread into the international literature in the 2nd and 3rd generations. By the 3rd generation, the publications more accurately reflected the international pattern of research output.

From the funding acknowledgements, the most striking observation is the high proportion of papers supported, at least partially, by the British Diabetic Association (BDA).

2.4.2 Key lessons learnt about how to conduct such an exercise

The application of a range of standard bibliometric techniques produced a considerable body of information. This type of analysis, however, becomes very time-consuming when the number of papers and citations becomes as large as happened here.

3 CATEGORISATION OF CITATIONS

3.1 Introduction

Chapter 2 presented results of a bibliometric analysis of the papers in this preliminary study. This Chapter is concerned with a detailed analysis of the citations. We start by describing a review of previous literature in this field. This was drawn on to devise an assessment/appraisal template to be applied to the 2nd generation papers. The template captured information relating to each occasion the 1st generation paper was cited in the paper, including its location and whether it was applying, supporting, developing or refuting a method or concept in the cited article or whether it was merely noting or reviewing it. The class of the article, whether it was basic or clinical, was recorded and an assessment made of the overall importance of the cited article to the citing article.

3.2 Literature review: Categorising citations

This review highlights both items and procedures included in previous studies of citations, and their implications for our own citation categorisation. The most thorough previous general reviews have been those by Case and Higgins (2000), McCain and Turner (1989), Cronin (1984) and Small (1982). Small, and later Vinkler (1988), also usefully compared the various classification schemes previously used. Small (1982, p.300) suggested that 'by and large, the development of citation classification schemes has not been a cumulative endeavour'. He went on to claim, however, that on closer inspection 'there are striking parallels between the schemes proposed, even though the terminology used to describe the categories can differ'. These similarities enabled him to devise his comparative tables. Work in this field is viewed as being quite complex with Case and Higgins (2000) referring to 'epistemological and methodological problems' (p.636) and Cronin (1984) suggesting 'the process is not amenable to scrutiny' (p.57).

The main purpose of our study, to assess the impact of basic or early clinical research through several generations of citations, is rather different from those of most earlier studies. Therefore, although we drew lessons from previous work, we needed to use it in a way that helped with our specific aim to use citation content/context analysis, in combination with citations tracing through several generations, to assess the payback from basic/early clinical research.

Following this introduction to the literature review, there are sub-sections on:

- the aims of analysing citations;
- procedural issues: who undertakes the classification of citations and to which citations is it applied?;

- what to include in a classification: motives/reasons for citing and strength of impact on the citing paper;
- the location and frequency of citations within an article;
- combining the factors.

3.2.1 The aims of analysing citations

The wider analysis of citations is undertaken for many reasons. Here we identify some of those most relevant for our project. Small (1982) concluded that citation context analysis is pursued by two groups: 'those concerned with information retrieval, and those concerned with the sociology of science and citation analysis' (p.308). The first stream of work, according to Small, started with Lipetz (1965) who wanted to improve information retrieval from citation indexes by devising a scheme of categories which an indexer would apply to each reference in a citing document. This stream of work proved to be impractical although Small (1982) and Cronin (1984) described various attempts to build on Lipetz's work: Finney (1979); Duncan *et al*, (1981). The importance of this work for us is that it describes attempts to move towards more routine approaches that we also see as being our ultimate goal.

The stream of work associated with the sociology of science started, according to Small, with a study by Moravcsik and Murugesan (1975). This has been the more productive stream. Moravcsik and Murugesan (1979) applied their previously developed method of classifying citations in scientific papers (1975) to studying two scientific revolutions in theoretical physics. Although concerned, like our project, with the development of science, their main interest was to analyse citations in order to understand 'paradigm changes' (Kuhn, 1970) rather than trace the impact of a certain body of work. Various other studies in this stream were motivated by a concern to assess the adequacy of citation counts as a measure of scientific impact (Small, 1982). Perhaps the analysis nearest to our concerns comes from studies that attempt to examine how citation patterns to certain highly cited papers might have varied over time (Cozzens, 1985; Oppenheim and Renn 1978; McCain and Turner, 1989). The latter took 11 highly cited papers in molecular genetics written between 1978-80. These 11 received a total of 3100 citations. Although for each of the 11 papers, they analysed only 10 citations in articles published in years one and two, and 10 in years six and seven, their approach was important and is further analysed in a later section.

Attempts to explain the nature of citations as either 'concept symbols' (Small, 1978) or as a process of persuasion (Gilbert, 1977) are important background papers for our study. Gilbert listed various reasons why a paper might be cited by an author as part of his attempt to persuade

readers of the importance etc of his own article. The author might 'cite the "important and correct" papers' (Gilbert, 1977, p.116) to add credence to his paper and might cite 'erroneous' papers to build up the importance of his advances. Gilbert noted, however, 'that respected papers may be cited to shine in their reflected glory even if they do not seem closely related to the substantive content of the report' (Gilbert, 1977, p.116). Gilbert's analysis has gained considerable support (see, for example, Brooks, 1985; Allen *et al*, 1994; and Case and Higgins, 2000) but highlights the difficulties in developing a template in a way that will accurately gather information. These issues are part of a wider, ongoing debate about the need for a theory of citations (see, for example, Cronin, 1984; Leydesdorff, 1998).

As noted in the introduction to the report, our proposed use of citations to track the impact of basic/early clinical research, through from the original papers to later work, is an approach that a recent major review of citations suggested was potentially useful but undeveloped (Kostoff, 1988). Gläser and Laudel (2001) recently claimed that 'research into establishing a typology of citations has ceased, and today we simply do not know how affirmative, rejective, perfunctory citations and the like are distributed. Consequently the knowledge accumulated by citation context analysis is not applied in 'normal' citation count-based studies' (p. 429). They also advocated the combination of quantitative and qualitative approaches.

3.2.2 Procedural issues: who classifies which citations?

Here there are two overlapping dimensions by which previous studies have varied:

- i. Who applies the categorisation to the citations being analysed;
- ii. To which citations the classification is applied?

i) Who undertakes the classification?

There are probably two options as to who should apply the categorisation to the citations being analysed:

- external reviewers/researchers;
- the authors who made the citation – either through completion of a questionnaire, or by interview.

In a series of early studies the reviewers each devised a categorisation of citations and applied it to a list of citations (Moravcsik and Murugesan, 1975; Chubin and Moitra, 1975; Spiegel-Rosing, 1977). Several teams stressed the need for specialist knowledge of the field covered by the papers, or at least a scientific background, to be able to undertake this classification and

Moravcsik and Murugesan (1975) also highlighted the desirability of testing reliability by having more than one reviewer and then comparing results.

There are obvious restrictions on the issues that an external reviewer can address. Therefore, Prabha (1983), in a study of references given in a source paper, directly sent a questionnaire to authors to explore issues such as how many of the sources had the citer read and how many of the sources were viewed as essential. Other advocates of the use of questionnaires to authors thought that questionnaires 'bring us closer to an understanding of citation behaviour' (Case and Higgins, 2000, p.645). There are, however, weaknesses with using questionnaires and Chubin and Moitra (1975, p.426) claimed, 'both the candour and recall of authors may be lacking, however, rendering such data impressionistic, selective and self-serving'.

From a different perspective some authorities are critical of the use of questionnaires because they, too, are unlikely to discover the full range of author motives in giving citations. Cronin (1984) and Shadish *et al*, (1995) advocated the use of in-depth interviews with authors and at least one such study (White and Wang, 1997) has been conducted. As Case and Higgins note (2000, p.645), however, in-depth interviewing, as with questionnaires, 'may suffer from problems of retrospective reasoning, recall and lack of honesty by respondents'. Furthermore, qualitative interviewing requires a great deal of time and effort. A further problem facing our study is the length of time since authors wrote their articles. This is particularly a problem compared to White and Wang's study, which was conducted at about the time the authors wrote their articles.

ii) To which citations is the classification applied?

There have been three main approaches to identifying the citations to which the categorisation should be applied:

- all the references *in* a source paper(s);
- for many different authors, a selected reference they have made in one of their articles;
- all (or a large selection) of the citations *to* the source paper(s) of interest to the study.

In the first approach, source papers (equivalent to our 1st generation) are identified through various approaches and then the classification is applied to each reference *in* each paper. The second option is more complex. It covers the work of Shadish *et al*, (1995) who identified a series of source papers and then randomly selected one reference from each article and sent the author of the article a detailed questionnaire about that one reference. We have put Bonzi and Snyder (1991) into this category, but they specifically examined self-citations. The third option, concentrating on the citations *to* the source article, is the one that our study will adopt: these

citations to the source paper occur in what we have defined as 2nd generation papers. Sometimes however, as with the McCain and Turner paper noted earlier, a selection of citations to the original source paper are taken rather than all of them.

Table 5 below attempts to classify previous studies according to the options in the two sub-sections above. It should be noted that at least one study, Chubin and Moitra (1975), started by examining the references in their source (1st generation) papers, and then analysed the citations to these papers. This study therefore appears in two boxes in the table.

Table 5: Who undertakes the classification and to what is it applied

To which citations is the classification applied	Who applies the classification	
	A External reviewers	B Authors who make the citation – via questionnaire or interviews
1. All the references in a source paper(s)	Lipetz (1965) Moravcsik & Murugesan (1975) Chubin & Moitra (1975) Spiegel-Rosing (1977)	Prabha (1982) Brooks (1985) Vinkler (1987) Cano (1989) White & Wang (1997)
2. For many authors a selected citation they have made		Bonzi and Snyder (1991) Shadish <i>et al</i> , (1995)
3. All (or a large selection of) citations to the source paper	Chubin & Moitra (1975) Cole (1977) Oppenheim & Renn (1978) Cozzens (1985) McCain & Turner (1989) Maricic <i>et al</i> , (1997)	Case & Higgins (2000)

3.2.3 What to include in a classification: motives and strength of impact

Some authorities in the field have produced lists of why citations are given. Garfield (1965) produced a list of 15 reasons and thought they were still valid 30 years later (Garfield, 1996). Lists such as this were designed as a result of the experience of working on, and thinking about, citations rather than, as in many of the empirical exercises described here, being developed for a specific empirical exercise. Shadish *et al*, (1995) decided, however, that it might not be appropriate to use an entirely externally generated list of reasons for citing and attempted to supplement the lists of previous authorities by systematically asking their own colleagues what had motivated them in citation behaviour. Similarly White and Wang (1997), in their direct interviews with authors about the reasons behind their citation lists, allowed the reasons to emerge from the authors.

We conclude that there is now so much previous research on which to build and sufficient externally generated lists of reasons for citing that it is unnecessary to take the exercise one stage further back and ask authors to help devise yet another new list. This conclusion is supported by the comments of Case and Higgins (2000) who, in their empirical study, built on the work of Shadish *et al*, (1995) and like them found that when they operationalised their research instrument, 'no other distinctive reasons for citing a document were uncovered in the open-ended responses' (p.639).

In the previous studies there are various dimensions, and combinations, of the question about what to incorporate in a classification of citations. These dimensions include:

- whether the focus is on motivations/reasons for citing, or on strength of impact the cited work has made on the paper citing it, or on a combination;
- the importance of including the questions that authors would be able to answer even though external reviewers could not;
- whether single or multiple motivations are linked to each citation.

Given this range of dimensions on which previous studies have varied, the following account does not attempt to provide a neat classification but instead provides a broad framework for considering key contributions. Some researchers, for example Spiegel-Rosing (1977), produced a list of motives for citing and attempt to classify all the references into one category or another. In this case it was not very satisfactory because of her 13 reasons, 80% were fitted into just one category. Other studies have allowed the external reviewer, or the citing author, to identify several motives behind each citation.

Moravcsik and Murugesan's 1975 study is seen as a key contribution. The categorisation they devised and then applied had four elements and in each one they classified each citation as belonging to one of two alternatives (or neither):

- Conceptual, Operational, neither
- Organic, Perfunctory, neither
- Evolutionary, Juxtapositional, neither
- Confirmative, Negational, neither.

Several authors started with the above classification and added variations. They include Swales (1986, reported in Ungern-Sternberg, 2000), and Chubin and Moitra (1975) who put references into one of 6 categories and attempted to add some measure of strength. Cano (1989) went back to the original four pairs of items and used them, but not only as a series of alternatives. This

resulted in some authors making clear that both options in some pairs, for example conceptual and operational, were motivations for some citations.

As noted earlier, Small (1982, pp.302-3) and Vinkler (1998, pp.133-4) attempted to show the similarities between a range of classifications previously devised by other authors. Small took the work of Lipetz as the 'Procrustean bed to which the others can be related' (1982, p.300). Having attempted to show the comparisons he then devised the following list of items, which are not seen as mutually exclusive (p.304):

- Refuted (negative)
- Noted only (perfunctory)
- Reviewed (compared)
- Applied (used)
- Supported (substantiated)

When questionnaires (or even interviews) with authors are used, this increases the number of motivations that may be enquired about. Vinkler (1987) made a distinction between 'professional' motivations and 'connectional' ones. The former are to do with the traditional items used in citation classifications such as that the citation may be: affirmative, comparative, perfunctory and/or negative etc. The 'connectional' motivations are to do with the social relationship between the citing author and the author of the paper cited. Others have used different terms for this category, but clearly most information about such motives can be gathered only directly from the authors, although some, such as whether the citing paper comes from the same institution as the cited paper, could be recorded by external reviewers.

Brooks (1985) asked authors to assess their motives for giving each reference along seven scales (although he also allowed authors to add in their own motives). Shadish *et al*, (1995), and the replicatory study by Case and Higgins (2000), used a list of about 28 possible reasons for citing. Authors were asked to plump for which was the most important, but then also asked to apply a five point Likert scale to each of the 28 reasons. Shadish *et al*, (1995) also asked a series of 'proximity' questions similar to Vinkler's connectional questions. They covered issues such as 'Have you ever spoken directly or by phone with the author of this citation?'

We have already noted that Prabha (1983) was one of the first to send questionnaires to citing authors. This enabled him not only to ask questions about whether the citing authors consulted the cited paper whilst actually writing their papers, but also to ask a question about the strength

of impact. For each paper cited, Prabha asked whether the author drew heavily, moderately or peripherally on the source.

This type of analysis would seem to be near to what we are attempting to identify in our project. Cano (1989) asked authors to state which of the eight items, derived as we have seen from Moravcsik and Murugesan's four pairs, was a motive behind their decision to include a reference. The most frequent of the eight categories was 'perfunctory', which was seen by Small (1985) as being equivalent to his term, 'noted only'. Cano then asked authors to judge the utility content of each reference they had cited. They were asked to rate on a four-point scale the level to which each reference had contributed to the production of the technical paper. The utility content of a reference was therefore defined as 'that which makes a reference indispensable in the production of novel information' (p.285). The four points were: peripheral, moderate, heavy and essential. The perfunctory references were generally seen to have low utility on this four-point scale.

3.2.4 The location and frequency of citations within an article

Cano (1989) recorded the position of each citation in terms of its distance (in fraction terms) from the start of the article. This location question was recorded along with more detailed motivational/strength of impact data noted in the previous section. Other authors (especially McCain and Turner, 1989; and Maricic *et al*, 1997) gave more emphasis to the location of a citation in terms of the section of an article in which it appeared. This, they correctly claimed, was a more objective factor than distance from the start. They used categories that could be applied in a fairly standard way, such as: introduction, methods, results, conclusion/discussion and McCain and Turner (1989) assumed that a citation in a methods section of a research paper was worth more than one in the introduction of a paper, or in a review article.

Information about the number of times an article was cited in the citing article was recorded in various studies (Oppenheim and Renn, 1978; McCain and Turner, 1989), with the former examining 978 papers that in 1974-5 cited some old highly cited papers in physics and physical chemistry. They found on average each paper was cited 1.13 times by each citing article. They noted Chubin and Moitra's claim that on average each cited paper they examined was cited 1.05 times by the citing article. They therefore concluded: 'our results indicate that there is a rule that states that each cited paper is referred to on average 1.05-1.15 times in every paper that cites it and which may be valid over much of science' (Oppenheim and Renn, 1978, p. 230).

3.2.5 Combining factors

In the preceding sections various items of information have been noted as probably worth recording, and at least one study (McCain and Turner, 1989) showed how various items could be brought together. Their Utility Index ‘combined citation occurrence counts, citation location, and citation context in a single measure of “perceived usefulness”’ (p. 149). Their approach, however, was rather mechanistic, and as noted above, made assumptions about the score to give a citation depending on, for example, predetermined ratings for different locations, rather than judging each individual citation. Furthermore, in their analysis self-citations were automatically given a lower score and there was a formula applied to cover, on a diminishing scale, the score for each time an article was cited in the citing article.

The various items discussed in the literature review informed our analysis of how to develop a template mechanism that combined factors so as to help us identify the papers in the 2nd generation that viewed the 1st generation papers (ie source papers) as being important. It was intended that analysis of the 3rd generation would concentrate on those papers that cited 2nd generation papers where the citation to the source paper was seen as being important.

3.3 Methods

3.3.1 Devising the template

The literature review above (section 3.2) informed the development of the template to be used for classifying each citation. The template was intended to capture information on all of the citations to 1st generation source papers (A1-A29). The template differed from the mechanistic approaches of some previous studies. (eg McCain and Turner, 1989). Our aim was to keep the template to one page and have clear definitions of the terms used. It was clear however, that in relation to motives two versions of the template were needed; one for articles and notes and one for reviews. The template covered:

- the location of each citation occasion (eg, for articles: Introduction, Materials & Methods, Results, and Discussion; for reviews: Introduction, Discussion and Elsewhere)
- the motives, or reasons, for citing the paper (ie, for articles: Develop, Support, Apply, Refute, Note/Review only; for reviews: Support, Refute, Note/Review only—it is important to record that for articles the category ‘Develop’ was added to list from Small (1982) described above);
- an overall assessment of the importance of the cited paper to the citing paper. This overall assessment was based on the reviewers’ application of a scale with a definition for each of four levels that was very similar to that used by Cano (1989) and again described above (Peripheral, Moderate, Considerable and Essential);

- information about whether the paper described basic or clinical research. The attempt here was, instead of just relying on the classification of journals, to undertake the analysis at the level of individual papers as had been previously suggested by Mason *et al* (2001).

Some members of the Advisory Group applied the first draft of the template to a limited selection of 10 papers (articles and reviews) and their comments informed changes. The final version of the template, including the definition of terms, is shown in Appendices 2-4 along with the instructions on how to use it.

3.3.2 The Pilot study

Two research assistants were employed to photocopy the relevant articles on which we planned to apply the template. Only articles, notes and reviews in English were copied and review articles which were longer than 70 pages were not included in the study. Iain Frame, the two research assistants, and Philip Green and Stephen Hanney all applied the template to a selection of 50 2nd generation papers. Assessors 1-5 were regarded as non-experts in the field of diabetes although assessors 1 and 4 had a background in natural sciences. Subsequently, we also asked a scientist who was involved in basic diabetes research, to apply the template to the same 50 papers.

The 50 papers from the 2nd generation that were used in the pilot consisted of 41 original articles and 9 reviews. Of these 50 papers, 46 cited just one 1st generation paper, two of them cited three of our 1st generation papers, and another two both cited five 1st generation papers; thus giving a total of 62 examples of citing papers. The intention was that all six assessors should apply the template to each of these 62 examples. The results were entered into a database and the entries checked by a second person. Statistical analysis of inter-rater reliability was undertaken. The results of this pilot study are shown in section 3.4.

3.3.3 Application of the template

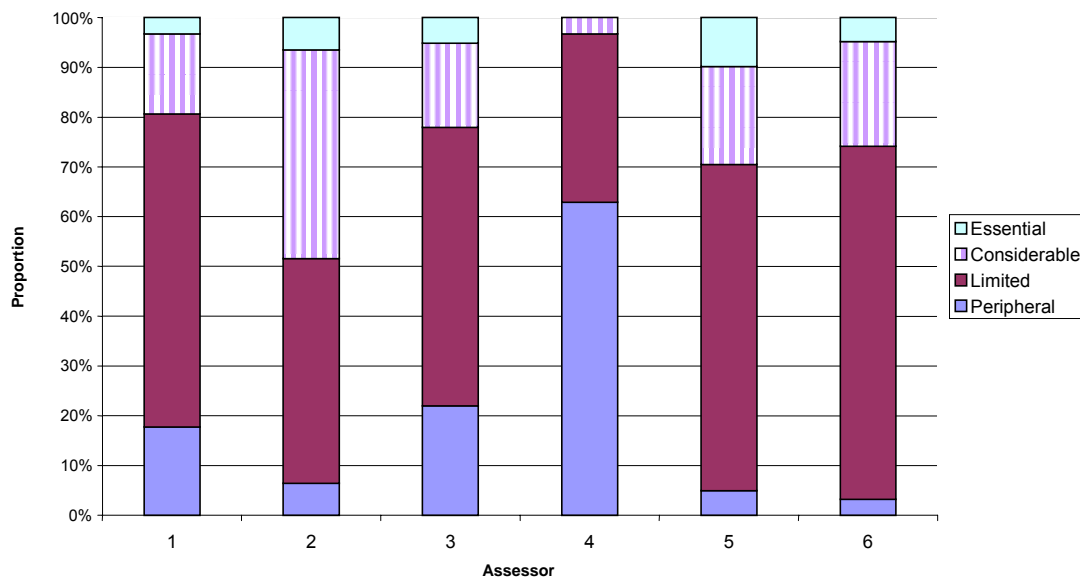
The template was also applied by one assessor (IAF) to the photocopies of most of the remaining 2nd generation papers. On-screen forms allowed direct entry of the collected data. The results from this main phase are described in section 3.5.

3.4 Results and lessons from the Pilot study

3.4.1 Importance of the cited article to the citing article

There is a reasonable similarity in the broad pattern of the results from the assessors analysing the 62 examples using the 4 categories (Peripheral, Limited, Considerable or Essential) to describe the overall importance of the cited article to the citing article. These are shown on Figure 11.

Figure 11: Overall proportions of categories of importance of cited article to citing article as rated by six assessors



There is something of a spectrum which is particularly clear if Peripheral and Limited are classed together as low importance and Considerable and Essential are classed together as high importance. Assessor 4 is at one end of the spectrum: he regarded the vast majority of the cited papers as either Peripheral or Limited and categorised no papers as Essential. Assessor 2 is at the other end of the spectrum, with an almost exactly even split between high and low importance. The dividing line for the each of the remaining four assessors fell within a few percentage points of having one quarter as high importance and the rest low. These four in the middle include assessor 1, the lead researcher and co-deviser of the template, assessor 5, the other co-deviser of the template, and assessor 6, the diabetes researcher. In the subsequent analysis, therefore, particular attention was given to the level of agreement between these three assessors.

In Table 6a the overall level of agreement as to the importance of the cited article is shown, and then the same when the four categories are collapsed in to the two: high and low importance. For five of the 62 examples, however, data were not available for one or more of the assessors. Therefore, the analysis relates to the remaining 57 examples of 2nd generation papers citing 1st

generation papers. An analysis of inter-rater reliability was carried out using the kappa coefficient (Stata Press, 1999). The kappa statistic is scaled where a value 0 denotes agreement that would be observed by chance, and 1 denotes perfect agreement. To interpret the kappa statistic six terms are used: poor (which is a score of less than zero); slight (0.00-0.20); fair (0.21-0.40); moderate (0.41-0.60); substantial (0.61-0.80) and almost perfect (0.81-100). The kappa statistic for each column is shown on Table 6a. Given the spectrum shown in Figure 11, it is perhaps not surprising that the level of agreement is classified as only fair, even when the categories are collapsed. Table 6b reveals that the agreement is stronger for the three assessors: 1, 5 and 6.

Table 6a: Agreement on importance of cited paper to citing paper: all 6 assessors

No of assessors agreed	No of agreed ratings (%) using all four categories of importance	No of agreed ratings (%) using High and Low importance categories
All 6	3/57 (5%)	24/57 (42%)
5	14/57 (25%)	19/57 (33%)
4	23/57 (40%)	10/57 (18%)
3	13/57 (23%)	4/57 (7%)
2	4/57 (7%)	N/A
Kappa	Slight	Fair

Table 6b: Agreement on importance of cited paper to citing paper: assessors 1,5 and 6

No of assessors agreed	No of agreed ratings (%) using all four categories of importance	No of agreed ratings (%) using High and Low importance categories
All 3	29/57 (51%)	40/57 (70%)
2	25/57 (44%)	17/57 (30%)
No agreement	3/57 (5%)	N/A
Kappa	Fair	Moderate

3.4.2 Categorising citations

For each cited paper, the assessors were asked to note each occasion a citation was given to a citing paper. As noted above, for the detailed analysis it was important to have information from each assessor on a common basis. All six assessors were in total agreement on the number of citations for 42 of the 57 papers (74%), and for the three assessors the figure rose to 49 (89%). It is difficult to give a definitive explanation as to why the assessors sometimes differed on the number of citation occasions. [In some cases, the poor quality of the photocopy may have

contributed.] However, there was no set pattern. Each assessor was capable of missing citation occasions compared to the rest of the group. For each occasion, the location of the relevant citation was recorded as: Introduction, Materials & Methods, Results, or Discussion in the case of notes and articles; or Introduction, Elsewhere, or Conclusion in the case of reviews.

Quite a few of the 57 examples of all assessors analysing 1st generation papers involved the paper being cited on more than one occasion in the same 2nd generation paper. This gives us a total of over 100 citation occasions. There were various “mismatches” where one or more of the assessors had recorded fewer citations in different locations. But there were 70 occasions/locations that matched between all 6 assessors and therefore the analysis concentrated on them. The complete analysis of the assessors’ scores was complex, involving the individual analysis of separate databases. The statistical analyses referred to below included all six assessors and data for the three assessors in several forms (but, to facilitate comparisons, the analysis of the responses from the three assessors was limited to the same 70 occasions where data were available for all six assessors).

Each of the 70 examples was examined to see whether, using the definitions given in Appendix 4, the citation had been categorised as Apply, Support, Develop, Refute, or Note/Review only. Again inter-assessor reliability was assessed using the kappa coefficient measure. The level of agreement between the assessors varied greatly over the five categories: whereas for Apply the agreement was almost perfect, at the other extreme it was only slight for Develop. For Support, Refute and Note/Review only it was fair (Table 7), but for each of these categories it was higher when just the three assessors were considered. Since assessor 1(IAF) was to score all available papers, his scores are compared separately with assessor 5 and with assessor 6. The picture is complex. For both of these latter two sets of comparisons there is one example where the kappa classification is lower than for the comparisons between all six assessors, but, overall, both sets show higher kappa classifications for level of agreement than those found for all assessors.

In all four sets the lowest level of agreement is in the Develop category, and this probably had some spillover effect into the comparatively low agreement for Support, which is the nearest alternative category into which citation occasions could have been classified. By contrast, the level of agreement for the Apply category is very high and many of the citations located in the Materials & Methods section of an article were in this category.

Table 7: Levels of agreement between assessors using kappa classifications

Category	Level of agreement between all six assessors	Level of agreement: assessors 1,5 and 6	Level of agreement: assessor 1 and assessor 5	Level of agreement: assessor 1 and assessor 6
Apply	Almost perfect	Almost perfect	Almost perfect	Substantial
Support	Fair	Moderate	Fair	Moderate
Develop	Slight	Slight	Poor	Slight
Refute	Fair	Substantial	Almost perfect	Substantial
Note/Review only	Fair	Moderate	Substantial	Moderate

3.4.3 Lessons from the pilots

The analysis above suggests that it was reasonable, within the limited resources available for the pilot study, for assessor 1 to conduct the application of the template to the remaining papers. There are, however, lessons from the pilots that suggest further preliminary work should be conducted to increase common understanding of the various categorisations applied, and thus improve inter-assessor reliability, before we move on to any large-scale study. These could take various forms.

First, attention needs to be given to the definitions of the various categories to enhance agreed interpretation. This could be assisted by an attempt to state how the various types of categorisations should be combined. There were situations, for example, where some assessors were categorising occasions as being Develop or Support and yet describing the cited paper as having only Peripheral importance to the cited paper. These would appear to be unlikely combinations. A somewhat more mechanistic approach towards identifying the possible combinations most likely to be accurate should help address this problem, as should clearer definitions.

Second, it seems that the greatest difficulty in terms of categorising the citation occasions came with the category that the team added to the existing lists, namely Develop. Perhaps this category should not be used in future. Finally, there is clearly a need for all of those applying the template to meet after applying it on a few occasions and work through together their various interpretations of the definitions and attempt to reach a consensus about how to apply them.

3.5 Results of the expanded template application to all 2nd generation articles

3.5.1 Overall importance of the cited article to the citing article

In total, 623 out of 799 citations to the 29 1st generation papers were scored by assessor 1 using the template. The results for the assessment of overall importance of the cited article to the citing article are shown in Table 8. For the majority of articles (351;56%) the 1st generation papers that they cited were assessed to be of Limited overall importance, and for a further 219 (35%) the 1st generation papers were classified as being of Peripheral overall importance. Only for 5 (1%) papers were 1st generation articles classified as being of Essential importance, with a further 48 (8%) classified as Considerable.

The median year of publication of the 2nd generation citing articles (and range) is also shown on Table 8 and suggests that the importance of the cited to the citing article becomes less with time. This finding is important in its own right. It also helps to explain why the proportion of papers in the two categories (Considerable and Essential) that can be collapsed to produce the high importance category is much lower in the overall analysis than it was at the pilot stage for assessor 1. As described above, the pilot stage concentrated on early papers from the 2nd generation where, as it has turned out, there was most likelihood of finding papers of high importance.

Table 8: Scores of overall importance of the cited article to the citing article with median publication year and range of citing articles

Overall Importance	Number (%)	Median publication year (range)
Peripheral	219 (35)	1988 (1981-2001)
Limited	351 (56)	1986 (1981-1999)
Considerable	48 (8)	1984 (1982-1995)
Essential	5 (1)	1983 (1982-1987)

Notwithstanding that the numbers of citations had become so large that it was impossible to use citation categorisation to trace significant impact through several generations, it was thought important to attempt to explore what use could be made of the effort that had gone into categorising the 2nd generation. (A very small amount of categorisation of the 3rd generation was possible, but this is not reported here). Table 9 shows, for each 1st generation paper, the numbers (and proportions) of 2nd generation papers that were classified as being in each of the four categories of importance.

Data from Table 9 can be analysed in various ways. At an overall level statistical analysis was undertaken to see whether there was a correlation between papers that were highly cited and the proportion of those 2nd generation papers in which the 1st generation paper was viewed as being of high importance to the citing paper. For this analysis the two collapsed categories of high and low importance were used. A χ^2 test was performed in order to establish whether there was a relationship between the number of times an article was cited and the proportion of citations of high importance. This showed there was no significant difference across articles: $\chi^2_{28} = 31.5$, $p = 0.295$. There was no evidence of correlation between the proportion of high importance papers and the total number of citations $\rho = 0.098$. The analysis was repeated, with similar results, after the removal of self-citations and also of papers with first less than five citations, and then less than ten.

Table 9: Number of times article is cited by level of importance

Paper	Peripheral	Limited	Considerable	Essential	Total papers scored (and total number of citing papers etc)
A1	2 (33%)	3 (50%)	1 (17%)	0	6 (9)
A2	2 (22%)	4 (44%)	3 (33%)	0	9 (9)
A3	18 (44%)	20 (49%)	3 (7%)	0	41 (47)
A4	3 (33%)	6 (66%)	0 (0%)	0	9 (13)
A5	2 (33%)	3 (50%)	1 (17%)	0	6 (10)
A6	19 (48%)	20 (50%)	1 (2%)	0	40 (47)
A7	10 (33%)	17 (52%)	6 (18%)	0	33 (42)
A8	9 (28%)	19 (59%)	3 (8%)	1 (4%)	32 (44)
A9	1 (50%)	1 (50%)	0 (0%)	0	2 (3)
A10	2 (40%)	3 (60%)	0 (0%)	0	5 (6)
A11	20 (53%)	17 (45%)	1 (3%)	0	38 (46)
A12	0 (0%)	1 (100%)	0 (0%)	0	1 (1)
A13	14 (25%)	38 (69%)	3 (5%)	0	55 (76)
A14	0 (0%)	6 (100%)	0 (0%)	0	6 (8)
A15	4 (31%)	9 (69%)	0 (0%)	0	13 (18)
A16	6 (46%)	5 (38%)	2 (15%)	0	13 (14)
A17	0 (0%)	6 (100%)	0 (0%)	0	6 (8)
A18	21 (45%)	23 (49%)	3 (6%)	0	47 (52)
A19	1 (100%)	0 (0%)	0 (0%)	0	1 (1)
A20	12 (23%)	34 (64%)	5 (10%)	2 (4%)	53 (74)
A21	3 (50%)	2 (33%)	1 (17%)	0	6 (9)
A22	6 (27%)	15 (68%)	1 (5%)	0	22 (27)
A23	8 (32%)	15 (60%)	1 (4%)	1 (4%)	25 (32)
A24	12 (38%)	17 (53%)	2 (6%)	1 (3%)	32 (37)
A25	9 (60%)	6 (40%)	0 (0%)	0	15 (24)
A26	14 (48%)	12 (41%)	3 (10%)	0	29 (39)
A27	6 (19%)	20 (62%)	6 (19%)	0	32 (41)
A28	11 (41%)	15 (56%)	1 (4%)	0	27 (38)
A29	4 (21%)	14 (74%)	1 (5%)	0	19 (24)
Total	219 (35%)	351 (56%)	48 (8%)	5 (1%)	623 (799)

Of the 623 papers that were categorised 114 (18%) were self-citations. This is consistent with a study of over 45,000 publications in Norway in which Aksnes (2003) found that 17% of citations received by papers in clinical medicine were self-citations and for molecular biology and genetics the figure was 22%. Although self-citations are often viewed as unacceptable in evaluations, it is argued by Pichappan and Saravady (2002) that their role deserves further analysis. They list various reasons why authors give self-citations including factors, such as increasing the visibility of earlier work, that could be viewed as being important activities for our attempts to trace impact.

Further analysis was conducted of our data-set to see whether there was a difference between self-citations and non self-citations in terms of the proportion of 2nd generation papers where the 1st generation paper was viewed as being of high importance. There was a significantly higher proportion (20%) of self-cited articles considered to be of high importance than of non self-cited articles (6%) ($\chi^2_1 = 24.41$, $p < 0.001$). Again the tests were repeated, with similar results, after the removal of papers with first less than five citations and then less than ten.

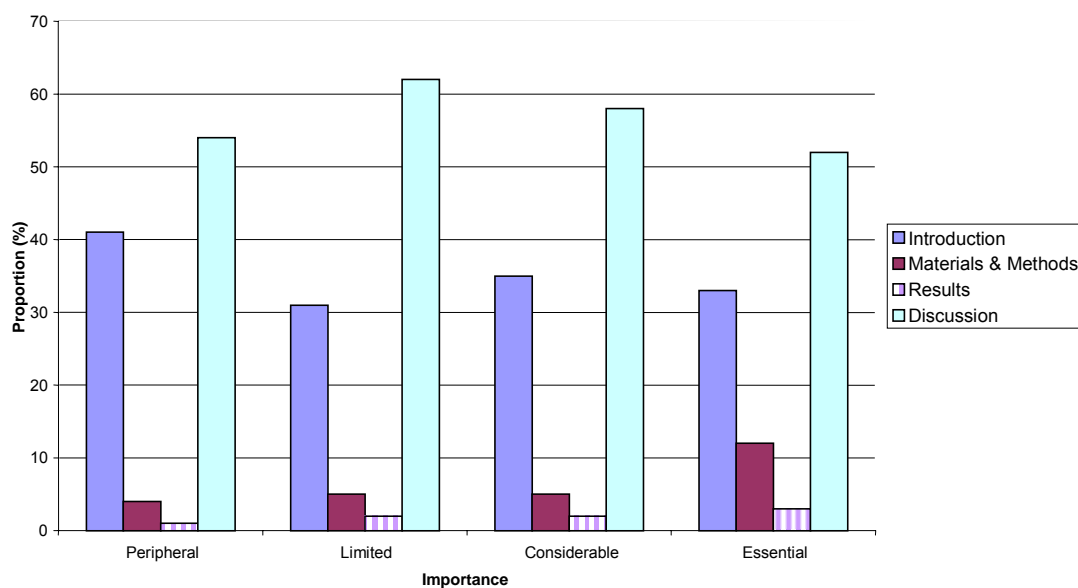
Extending the analysis of the figures on Table 9 to consider some individual 1st generation papers complements various points from the overall analysis. Paper A2 has the highest percentage of citations viewed as being of high importance: three out of nine (33%). Two of the three papers of high importance, however, are self-citations. Both A14 and A17 had all their citations classified as being of limited importance, which means they get a score of zero for papers of high importance, but also zero for papers of peripheral importance. Comparisons can also be made between the number of citations and proportion of them viewed as being peripheral by the citing paper. Looking at the four papers with most citations further illustrates the lack of a consistent pattern. Papers A13 and A20 have 25% and 23% of citing papers classified as viewing the importance of the paper as peripheral. This is well below the overall average shown on Table 9 of 35%. For the other two in the top four, A3 and A18, the figures are somewhat above the average at 44% and 45% respectively. This analysis will help inform the qualitative approaches described in the next chapter.

3.5.2 Importance of the cited article and location of the citation occasion

There were only five 2nd generation papers where the cited 1st generation paper was thought to be Essential, and two of the five classified the same paper as Essential. In each case, however, the 1st generation paper was cited on multiple occasions in the 2nd generation paper. Thus these five 2nd generation papers contained 36 citation occasions, and the location of each was separately recorded and categorised as is shown in Appendix 5.

Figure 12 shows the relationship between the importance of the cited article to the citing article and the location of the citation occasions. The figures for Essential show how the 36 examples were distributed between the four locations. By contrast, there were 219 citing articles in which the cited paper was classified as Peripheral. Many of these citing articles would have contained only a single citation occasion, although some could have contained several.

Figure 12: Location of citation vs Importance of citation



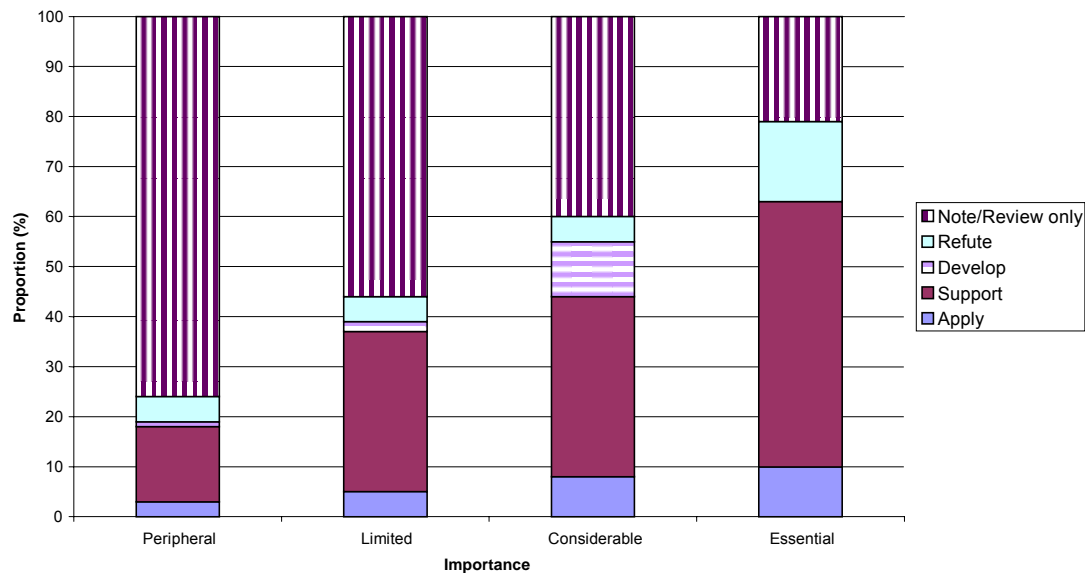
The overall pattern for each category of importance was the same, with most citation occasions being located in the Discussion, followed by the Introduction, Materials & Methods, and finally the Results. Nevertheless, there are some interesting details. Four out of the 36 citation occasions appearing in papers where the cited article was judged to be of Essential importance to the citing article occur in Materials & Methods sections. This means that, in percentage terms, the proportion of citation occasions appearing in the Materials & Methods section that are classified as being Essential are twice as high as for any other category of importance. This does not indicate, however, that citation occasions within a Materials & Methods location are more likely to be classed as Essential than anything else. Instead, it is more a reflection of the fact that, as shown in Appendix 5, 1st generation papers classified as Essential tend to have their multiple citations occurring in various parts of a citing paper, including Materials & Methods.

3.5.3 Overall importance of the cited article and classification of citation

The relationship between the overall importance of the cited article to the citing article and the classification of citation occasions is shown in Figure 13. Again the column for the articles

classified as Essential shows the categorisation proportions for the 36 citation occasions occurring in the five 2nd generation papers. Only papers classified as articles or notes were used in this analysis.

Figure 13: Classification of citation vs overall Importance of cited article

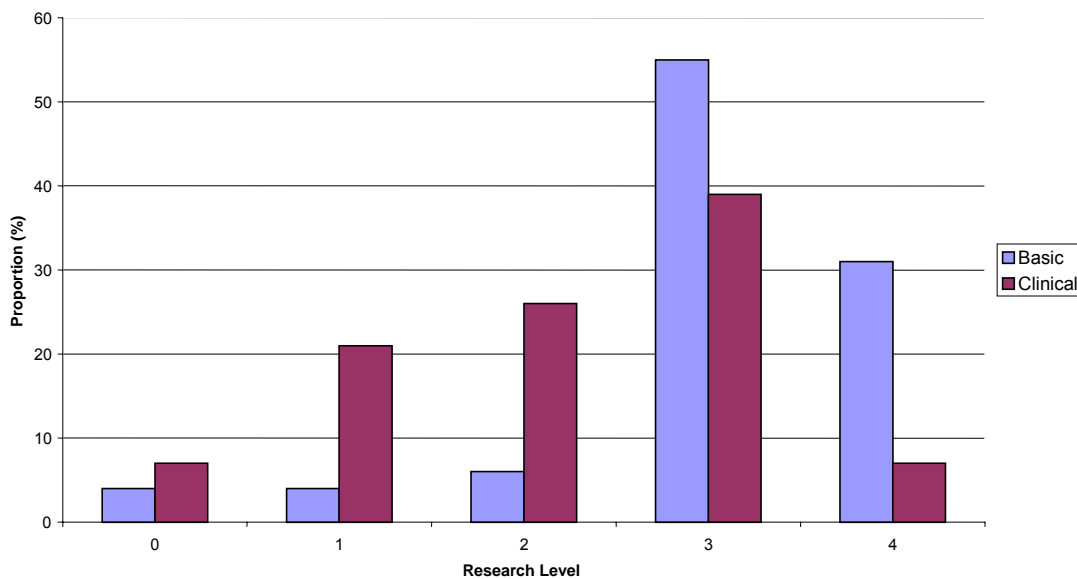


In all four levels of importance the most frequent classifications were Note/Review only and Support with relatively few citation occasions classified as Apply, Develop or Refute. The most striking trend in the classification of citations across the four categories is the shift from a majority of Note/Review only citation occasions in the Peripheral articles to a majority of citation occasions categorised as Support in the Essential articles. Almost 80% of citations in those articles deemed Peripheral were Note/Review only compared to just over 20% in those classified as Essential. Conversely, over 50% of citations in those papers categorised as being of Essential importance were classified as Support compared with less than 15% of citations in the Peripheral importance category. Given that the focus here is on citation occasions, it is understandable that there are some occasions within articles classed as Essential that are only described as Note/Review. It might be more surprising that there are as many as 15% of citation occasions in the Peripheral importance group that are classified under Support. As there are only five 2nd generation papers in the Essential group it is perhaps not too surprising that there are no occasions classified as being Develop. Nevertheless, this might cast further doubt on the usefulness of the Develop category.

3.5.4 Article class and Research Level

The concept of assigning a Research Level to a particular Journal has been described previously (Chapter 2). Individual papers within the 2nd generation were assigned to one of two categories, Basic or Clinical⁴. Comparisons of article class and Research Level are shown in Figure 14. Whilst the majority of papers describing basic research tend to be published in RL3 or RL4 journals, papers describing clinical research appear in RL1, RL2 and RL3 journals with, as expected, few appearing in RL4 journals. This finding broadly supports the validity of the calculation of Research Levels by CHI Research Inc although, as expected, indicates that it might not be valid in all cases. An additional complication is that these 2nd generation papers were published between 1981 and 2001, but the Research Level assigned to each journal is fixed and does not reflect possible changes in the character of the journal (Lewison and Paraje, 2003).

Figure 14: Research Level vs Article class for 2nd generation articles



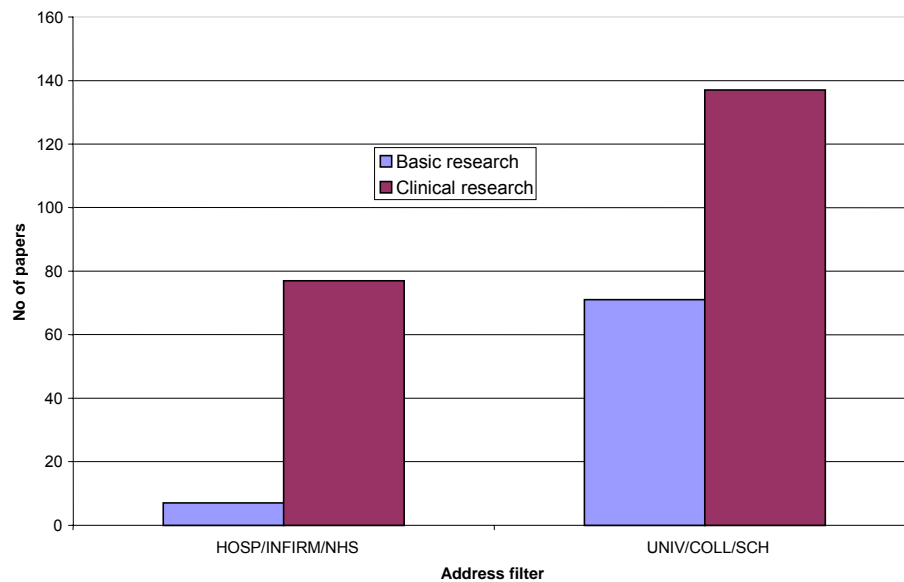
3.5.5 The research setting

Where they were available, the research settings of the 2nd generation papers were examined using two address field filters. The UNIV/COLL/SCH filter identified academic addresses, and HOSP/INF/NHS was used for NHS sites. Three hundred and thirty-nine papers had an academic address and 215 papers had an NHS address, but of these 131 had both. The level of research represented in the papers with just one type of setting is shown in Figure 15. There were 84

⁴ **Basic research-** theoretical or laboratory studies; **Clinical research-** relating to the observation and/or treatment of human subjects with a view to improving human health.

papers with just an NHS address, 77 (92%) of which were describing clinical research and 208 papers with just an academic address, 137 (66%) of which were describing clinical research.

Figure 15: Analysis of clinical and basic research settings



It has previously been assumed that research undertaken in a university setting is correlated with basic research whilst research undertaken in a hospital setting is correlated with clinical research (Grant *et al*, 2003). Based on the evidence from this small body of work in the diabetes field, this assumption could be questioned and further investigation is warranted, particularly of the work emanating from university-based research. For example, collaborative research between university-based hospitals and academic departments within the university could well be important.

3.6 Key findings

3.6.1 Key observations about the payback

The cited articles were judged to be of Considerable or Essential importance for only 9% of the 623 2nd generation papers that were categorised.

No significant relationship could be established between the number of times a paper was cited and the frequency of citations in which the cited article is classified as being of high importance. Self-citations, however, were significantly more likely than non self-citations to be classified as being of high importance.

It is quite striking that a few 2nd generation papers cited the 1st generation papers on multiple occasions and this was true for all 5 papers classified as Essential.

Classification of the type of research (basic or clinical) by paper compared to the classification of the journal by Research Level is broadly in agreement although, as expected, there are some differences.

3.6.2 Key lessons learnt about how to assess payback

Using the literature it was possible to build a template for categorising citations. Work is still required to further define the various categorisations used in the template in a sufficiently unambiguous way to improve inter-assessor reliability.

The considerable time required to apply the template, plus the lack of any overall pattern in terms of correlations between number and importance of citations, might point to the desirability of adopting a more selective approach aided by qualitative analysis. In any selective approach, however, it is likely that self-citations should feature.

4 QUALITATIVE ANALYSIS

4.1 Introduction

It was hoped that qualitative analysis would provide a fuller understanding of the processes by which research can influence subsequent studies and lead to wider benefits. Various qualitative approaches have proved useful when examining the impact that health services research has made on policy and practice, and on the career development of the researchers. These include: interviews with researchers, peers and other stakeholders; documentary analysis; and questionnaires to researchers (Buxton and Hanney, 1996; Buxton *et al*, 2000; Lavis *et al*, 2001; Hanney *et al*, 2003). In those studies it proved advantageous to start with a range of methods, and to be open to adopting extra assessment strategies as opportunities arose. This current study was intended to be a preliminary study and therefore it seemed particularly appropriate to test a range of methods and be flexible about adopting additional approaches.

We initially proposed to use three main methods for the qualitative analysis:

- i. questionnaires to co-authors of all 29 papers from 1981—the 1st generation papers;
- ii. development of critical pathways by selected experts/peers;
- iii. interviews with researchers, peers and other stakeholders.

Each of these is described in the following sections. In addition, Alberti, during a wide-ranging interview, raised the possibility of adopting an alternative strategy: examining the impact of a selection of key papers co-authored by him over a much longer period in the 1970s and 80s. This therefore eventually constituted an additional, fourth, approach:

- iv. analysis based on a selection of 10 key papers from Alberti.

It was possible to incorporate analysis of these 10 papers into the interviews. This supplementary list also, however, helped to shed further light on, and contextualise, comments made in the critical pathways and questionnaires about the original 1981 outputs. The discussion of the results of the original methods is therefore informed by this additional approach; its methodological significance is further reviewed in the final chapter. This qualitative chapter describes an approach involving triangulation of both methods and data sources. The emerging findings and results from earlier parts of the study were built into the later design of interview schedules, and various points were tested on the interviewees. Following the account of the three original methods, an outline of the application of all four approaches is given. There is then a combined account of the significant findings. Some key observations linking findings and methods are made at the end of the chapter.

4.2 Methods: original plan

4.2.1 Questionnaires to 1st generation co-authors

Alberti's research collaborations were widespread. We saw in section 2.3.3, for example, that the 1981 list includes papers from all four main categories of Research Level, ranging from clinical observation to basic research. This breadth was also reflected in the number of co-authors involved in the 1981 papers: a total of 74. For such a large group a questionnaire approach seemed more appropriate than interviewing. The questionnaire was designed to cover a range of issues (see Appendix 6). It started with questions about the impact of the co-authored paper on subsequent research, before moving on to ask about any impact on policy, practice, commercial developments, or medical training. Finally, two more general questions cover the attraction of overseas researchers to Newcastle and overall impressions of the body of work from 1981.

Alberti readily agreed that it would seem appropriate to seek the views of his co-authors. Home, as a member of the project's Advisory Group, distributed the questionnaires from Newcastle, but confidentiality was guaranteed and the pre-paid reply envelope was to the Wellcome Trust.

A questionnaire to the authors of the 2nd generation papers would have sought information on a range of issues that had been identified as important in our review of citation behaviour (Chapter 3). Previous attempts to ask about such issues had, however, been conducted within a relatively short time of publication. Most of the authors we would have been approaching would have written their articles over ten years prior to receiving this questionnaire. Therefore, we concluded it would be difficult for them to answer many of the questions in which we were most interested. Furthermore, the analysis of the importance of the cited work in the citing paper was revealing that the citations that were still being given in the 1990s tended not to be rated as being of great significance. Therefore, the very authors who might have a clearest view about the issues in which we were interested, were the ones for whom the questionnaire would have least salience.

4.2.2 Development of critical pathways by selected experts/peers

The traditional assessment of research quality is left in the hands of peers. It was clearly highly desirable for this study to obtain the opinions of peers/experts. This exercise, however, was concerned with tracing the payback from research, rather than making an assessment of quality. The complexity of unravelling the impact of research is such that quite often an 'insider account' is seen as valuable (Phoolcharoen, 2002; Milbank Memorial Fund/the Cochrane Collaboration, 2001). Therefore, it seemed appropriate to commission at least one of the peer analyses from Home who had been a colleague of Alberti's for many years, including in 1981.

The length of time over which the analysis was focusing meant that the most appropriate form of expert/peer review could be critical pathway analysis. Such analysis involves experts considering how lines or paths of research, and researcher careers, have evolved. Critical pathway analysis was undertaken as part of the Policy Unit's study replicating Comroe and Dripps, but in the context of working backwards proved very difficult to do (Grant *et al*, 2003). For our study the experts undertaking the critical pathway analysis were asked to consider, in particular, the impact of each of the 29 papers on subsequent scientific advances and on health policies, practice, commercial exploitation and medical education (see Appendix 7). The instructions were deliberately not too prescriptive and the critical pathway authors were asked to detail the methods they adopted. An allocation of time of two days per critical pathway was suggested.

4.2.3 Interviews with researchers, peers and other stakeholders

Four interviews were conducted, face-to-face or by phone, and the interviewees each covered a number of perspectives. Alberti not only gave his view on the impact of the body of work from 1981, but also put it into wider perspective and discussed additional possible methods. Robert Tattersall is now an historian of diabetes, but previously was a leading academic clinician working at the same time as Alberti. The Director of Research at Diabetes UK, Moira Murphy, was previously a research scientist for a pharmaceutical company and in that capacity sometimes worked with the Newcastle team in the 1980s. She was also a member of the Department of Health/MRC Research Advisory Committee on Diabetes for the National Service Framework for Diabetes. Sue Roberts, too, was a member of that committee and of its Service Organisation and Delivery Subgroup. She is a leading consultant diabetologist, and winner of the Hospital Doctor of the Year Award in 1994, who has worked for many years in the North East of England, near to Alberti's Newcastle centre.

4.3 Results of applying the methods

4.3.1 Questionnaires to 1st generation co-authors

In total 24 questionnaires were returned and one was jointly completed by two co-authors who still work together. Many of the authors had collaborated with Alberti on more than one paper, in one case on as many as seven. Therefore, for some projects there was information from more than one questionnaire, and there was only one of the 29 papers for which there was no questionnaire completed. The information on the questionnaires varied enormously, however, with some respondents saying they did not think they had anything useful to contribute, to others who addressed all the questions and helpfully supplied some supplementary information. In preparation for the interviews, the questionnaire material relevant for each publication was combined with that from the critical pathway and the record of the number of citations.

4.3.2 Development of critical pathways by selected experts/peers

A critical pathway was completed by Home and, as described in this section, its content broadly mirrored the suggestions set out in the guidelines contained in Appendix 7. It supplied information on every paper, although inevitably there was less detail about some of the papers where there had been little subsequent impact and/or the research was conducted primarily outside Newcastle or outside of the main areas of diabetes research. Home did not at this point have access to our detailed bibliometric and citation analysis on each paper. He started by adopting a formal assessment approach that covered a wide range of items for each paper. It soon became apparent to him that this would be too time consuming. Therefore, he reverted to asking three main questions of each paper:

- Was there a discernible relationship between the paper and the first author's career?
- Did a body of research or commercial outcomes flow from the paper?
- Was there a discernible clinical impact, either in general or through a guideline?

As a result of considering these three questions Home was able to make a free text comment on each paper and a summary impact statement. In some cases it proved possible to group two or three papers together and make some general comments about them, for example, papers A7, A22 and A24 were grouped together under the heading human insulin. About them Home commented: 'These were amongst the first papers on the pharmacokinetics and pharmacodynamics of human insulin...These papers, with a small number of others, were critical in providing the information that underpinned the introduction and early clinical trials of human insulin'.

Home also observed, however, that the literature:

'is but a small part of the flow and understanding in any scientific area, and this is even more true as one moves to clinical practice. ...Most papers are like a new vegetable thrown into a pot of soup that is already in an advanced state of preparation—they may have an influence on the taste to varying extent, but it is the whole that generates the aromas that might influence the cook's next action. Certainly the concept of any linear flow of activity is not applicable to this set of papers'.

The task of independently constructing a critical pathway proved impossible for a more basic scientist in this field who was not a clinical diabetologist, and was abandoned.

4.3.3 Interviews with researchers, peers and other stakeholders

Each interview took a somewhat different course, and the concentration was partially on feeding aspects and issues raised during previous interviews into later ones. The information from the questionnaires and critical pathways was used extensively, however, in structuring each interview. On many issues a reasonable degree of agreement was expressed, but there was marked divergence of views on a few issues. Alberti, too, took the view that generally science advanced by incremental steps and that it was difficult to pick out much that had made a real difference. He felt this was particularly true of the 1981 set of papers, whereas there were a few others during his career that perhaps had been of greater significance. These related, in particular, to two main areas. First, identifying new approaches to treatment for diabetic coma and for diabetics during surgery. Second, looking for new methods of treatment related to insulins. In both cases, Alberti, as noted in the introduction to the report, emphasised his approach of identifying clinical problems and then using basic science where necessary to help find a way forward.

4.3.4 Analysis based on a selection of 10 key papers from Alberti

As also noted above, Alberti felt that a list of 10 selected papers from the 1970s and 1980s might give a more appropriate picture of the payback from his work. He therefore subsequently selected 10 papers that, he felt, had had some impact on clinical practice. These are presented as Table 10, and numbered S1-S10. The citations for each paper are shown on the table. (In this way Table 10 combines the information that for the 1981 set of papers is contained in Table 1 and Figure 3). Although the citation figures for the additional set have not been reduced by focusing on articles, notes and reviews only, as was undertaken for the 1st generation papers (see section 2.2.1 above), it is clear that these papers had an average citation rate about three times as high as the figure of 28 for the 1981 set.

In the subsequent interviews, this additional list was also sent to interviewees prior to the interview and discussed. Tattersall, in preparing for his interview, drew on his knowledge of the history of diabetes to prepare what virtually amounted to a partial critical pathway for these 10 papers. To some extent, the introduction of an additional list of papers distracted from the original purpose of the exercise. However, it not only provided a fuller picture of Alberti's work, it also created a context for understanding and assessing comments in some of the questionnaires that initially appeared to make rather substantial claims for some of the 1981 papers, but did so by saying they were part of a wider stream of work.

Various previous attempts by Diabetes UK to consider analysis of research impact indicated that it would be difficult to limit the approach to a single year. Furthermore, in practice, the additional papers did not take the exercise so far away from the original intention as might appear. Had the plan described in the initial project proposal proved practical, then a three year period would have been studied as described in the introduction. In this event, papers S5-S7 would have been included in addition to the 1981 paper already in both lists: S4/A7. Moreover, paper S8 was already in our 2nd generation and S10 would have been a 2nd generation paper from the 1981-83 1st generation.

Table 10: KGMM Alberti ---10 Selected Publications from 1970s and 1980s Relevant for Clinical Practice

S1. Alberti KGMM, Hockaday TDR, Turner RC. Small doses of intramuscular insulin in the treatment of diabetic "coma". *Lancet* 1973; 2: 515-522. **Citations: 199**

S2. Pickup JC, Keen H, Parsons JA, Alberti KGMM. Continuous subcutaneous insulin infusion: an approach to achieving normoglycaemia. *British Medical Journal* 1978; i: 204-207. **Citations:308**

S3 Alberti KGMM, Thomas DJB. The management of diabetes during surgery. *British Journal of Anaesthesia* 1979; 51: 693-710. **Citations:70**

S4. Owens DR, Jones MK, Hayes TM, Heding LG, Alberti KGMM, Home PD, Burrin JM, Newcombe RG. Human insulin: study of safety and efficacy in man. *British Medical Journal* 1981; 282: 1264-1266. **Citations:47**

S5. Worth R, Home PD, Johnston DG, Anderson J, Ashworth Linda, Burrin Jacqueline M, Appleton D, Binder C, Alberti KGMM. Intensive attention improves glycaemic control in insulin dependent diabetes without further advantage from home blood glucose monitoring: results of a controlled trial. *British Medical Journal* 1982; 285: 1233-1240. **Citations:102**

S6. Francis AJ, Home PD, Hanning I, Alberti KGMM, Tunbridge WMG. Intermediate acting insulin given at bedtime: effect on blood glucose concentrations before and after breakfast. *British Medical Journal* 1983; 286: 1173-1176. **Citations:56**

S7 Husband DJ, Alberti KGMM, Julian DG. "Stress" hyperglycaemia during acute myocardial infarction: an indicator of pre-existing diabetes? *Lancet* 1983; ii: 179-181. **Citations:37**

S8. Elliott Martin J, Gill Geoffrey V, Home Philip D, Noy George A, Holden Michael P, Alberti K George MM. A comparison of two regimens for the management of diabetes during open-heart surgery. *Anesthesiology* 1984; 60: 364-368. **Citations:7**

S9. The Kroc Collaborative Study Group. Blood glucose control and the evolution of diabetic retinopathy and albuminuria: A preliminary multicenter trial. *New England Journal of Medicine* 1984; 311: 365-372. **Citations:59**

S10. Gray CS, Taylor R, French JM, Alberti KGMM, Venables GS, James OFW, Shaw DA, Cartledge NEF, Bates D. The prognostic value of stress hyperglycaemia and previously unrecognized diabetes in acute stroke. *Diabetic Medicine* 1987; 4: 237-240. **Citations:29**

4.4 Findings from applying the four methods

The combined sets of papers reveal examples where the impact is agreed to have been extensive, but also papers that fall into wide range of other situations. As noted by Tattersall, the advantage of focusing on the papers from one year was that they provided a realistic picture of the large amount of research from any team that does not really lead anywhere. There were, therefore, various advantages in having two lists. Given that the interviews drew upon the information from the other approaches, and that there is such a range of types of impacts from the research, it is appropriate for the analysis of the results of the four approaches to be merged into one section. Therefore, the section that follows is organised according to a number of categories of impact made by Alberti's research.

4.4.1 Examples of important basic research

We here describe several of the 1981 papers that report basic research regarded as being of some significance in contributing to the development of the scientific field. For these papers there is a mixture of both broad agreement but also some differences of emphasis about their exact contribution. Paper A24 (see Table 1, p8) describes a study, conducted at the National Institute for Medical Research, Mill Hill, which Alberti played a major role in supervising. It examined the importance, in diabetic dogs, of portal and peripheral infusions of insulin. The findings seem to have been subject to slightly different interpretations, but the study is widely viewed as being an important one and it was suggested that the area has continued to be of scientific interest. This is because it is still thought that research into portal insulin delivery and hepato-specific insulin could eventually have positive clinical dividends. The comments are consistent both with the number of citations (37) being above the average for the 1981 set, and with the half-life for the citations, ie the year by which half the total citations had appeared, being slightly longer than average (see Figure 4, p13).

The Newcastle based study of the metabolic effects of growth hormone, and the role of insulin in countering them (A18), is seen as quite an important basic physiological paper. It was cited 51 times which made it the third most highly cited paper in the 1981 list (see Figure 3, p13). However, in terms of the classification given to the those citations (see Table 9, p36), for an above average number of the papers citing A18 the importance of its contribution was rated as only Peripheral. This paper was one of the occasions on which there appeared to be some divergence between the citation count and the view expressed in the critical pathway, which was that the research impact was not great. The large number of citations in which the contribution was seen as Peripheral might help explain this; it might also be compatible with Alberti's view that some of the ketone body studies from 1981 were interesting but would not have had any impact on clinical practice.

4.4.2 Basic research of diverse significance

There was considerable variation in the extent to which a range of other basic studies, either conducted at Newcastle or for which Alberti's team did the measurement of metabolites, led onto subsequent research or were cited. Some of them did, however, contribute to the overall body of knowledge. For some of them the techniques in use in our study produced even more differing messages than those given in the above example.

Paper A3 describes basic research involving the metabolism of somatostatin and its analogues by rat liver. With 47 citations it was jointly the fourth most highly cited paper of the 1981 set and one of the co-authors described its importance to some specific later papers. Furthermore, checking the classification that had been given to these specific citations revealed that paper A3 was indeed thought, in our categorisation process, to have been of considerable importance to these citing authors. Tattersall saw it as an interesting paper that did not seem to have had many clinical or practical consequences, and the research area appeared to have been one that had fizzled out. Despite this, the half-life of the paper was also the fourth longest out of the 1981 set, but again an above average number of its citations were seen as Peripheral (see Table 9). This is compatible with the impression of the team member undertaking the scoring of the citations that many citations were referring to a method of measuring the rate of degradation using talc, and that this had not seemed a particularly important point. This contrasts with some of the previous attempts to categorise citations that were described in section 3.2.4. In these, any citations appearing in the methods section of a paper would automatically be given a higher score for location than other citations would receive, irrespective of the content. Our less mechanistic approach suggests citations in any location are best judged on their individual merits.

In contrast to the divergent evidence about paper A3, there was agreement between all the methods on the limited impact of paper A19. This study examined the liver metabolites in resting and exercising rats at 1 and 4 bar. It was cited just once and that was classified as being of only peripheral importance to the citing paper.

4.4.3 Research with major and rapid clinical impact

There is complete agreement between the various data sources for the qualitative study that paper A5, describing the management of diabetes during open heart surgery, made a significant and continuing impact on clinical practice. According to Tattersall, for example, it was 'very important'. It is also claimed that it made an impact on subsequent research in the field and that it identified a new problem. For our study, therefore, it is most interesting to note that the paper received only 10 citations, well below the average figure for the set. Partly this is explained by it

relating to improvements in the management of diabetes during open-heart surgery, which is a comparatively small clinical field. Nevertheless, had our study relied solely on citation analysis the significance of this paper would have been overlooked. A full understanding of its contribution comes from looking at Alberti's selected list of papers from the longer period.

Paper S3, from that list, examined the management of diabetes in surgery generally. It is seen by Tattersall as 'a very important paper' which provides 'a comprehensive review of the literature'. There were many diverse ideas around at the time about how best to manage diabetes during surgery. This review finished with suggested guidelines that had a considerable impact on practice by establishing that the glucose-potassium-insulin regimen first developed by Sodi-Pallares should be applied in this context. It also strongly argued for anaesthetists using near patient methods to measure glucose. Some clinicians adapted the idea and used their own delivery mechanism, but many applied the Alberti technique. Paper A5, therefore, can now be seen as applying, to the specific context of the long period involved in open heart surgery, the ideas developed by Alberti and Thomas a few years previously in paper S3. Similarly, paper S8 was seen as a development of A5 and one that tested a particular piece of equipment, the biostator, during open heart surgery. In addition to being on Alberti's selected list, S8 was, as noted above, one of our 2nd generation papers. Despite the importance of the stream of work, paper S8, too, had a low citation rate: the figure of 7 citations was well below that of most of the papers on the 1981 list, let alone those on the S list.

The paper by Alberti and colleagues on small doses of intramuscular insulin in the treatment of diabetic coma, S1, also enjoyed rapid and extensive application. According to Tattersall it was influential in Europe in the movement away from giving large doses of insulin. Roberts, for example, read it as a Senior House Officer and regarded it as very important – she still occasionally refers clinicians to it. Again the provision of quite detailed recommendations for treatment probably helped account for its considerable impact. Murphy contrasted the benefits from this situation with that for children, where it had taken a long time before equivalent guidelines had been produced. Furthermore, with 199 citations paper S1 has clearly had considerable impact on subsequent research.

The significance of both papers S1 and S3 lies in their clinically relevant recommendations that had immediate and long-lasting impact. Paper S1 represented quite a significant break from previous work, although Sönksen and colleagues were also exploring somewhat similar things at the same time (see, for example, Sönksen *et al*, 1972). In the case of S3, in particular, there was clearly much preceding work, but it was the ability to build on that work to make the clinically relevant recommendations that was so important.

4.4.4 Early clinical research of mixed immediate impact but major long-term importance

The studies considered here, A26, A8, and S2, were part of a collaboration between Newcastle and Guy's and perhaps come nearest to the model implied in the original application from the Wellcome Trust/Brunel. The consequences of the work described in these papers can be traced through various generations. They can be seen to have had considerable impact on subsequent research that is now, in turn, impacting on practice. The story is far from straight forward, however, and again it is not really possible to comprehend the full picture by limiting the analysis to the 1981 papers. Paper A26 describes a study showing that for provision of insulin before meal times a 'bolus' delivery was optimal. The study was of some importance in its own right and received 39 citations. It was, however, part of a much wider stream of work on continuous subcutaneous insulin infusion using a subcutaneous pump.

Paper S2 is one of the main papers describing the origin of using subcutaneous pumps and it resulted from Alberti's collaboration with Mill Hill, where John Parsons used such pumps for more basic research, and with colleagues at Guy's who were conducting research into diabetes. Paper S2 could perhaps best be described as early clinical research. This paper has been cited over 300 times and various comments about paper A26 were clearly related to the flow of subsequent research that resulted from the work described in a series of papers from the Guy's/Newcastle/Mill Hill team in the late 1970s and early 1980s. In particular, the pump addressed the issue of insulin delivery in a controlled way that mimicked insulin secretion in non-diabetics. Major research projects were undertaken to examine the clinical benefits that might come from the consequent improvements in blood glucose control.

The international study funded by the Kroc Foundation is seen as flowing from this stream of work. This was a prospective multicentre randomised trial to determine both the feasibility of maintaining improved blood glucose by using continuous subcutaneous infusion using pumps, and the effect of the improved control on diabetic microangiopathy and albuminuria. Newcastle provided the central laboratory for this Anglo-North American study, and Alberti was one of the authors of the key paper describing the project. This appears on the selected list as S9, but the explicit reference in this paper to the work on subcutaneous pumps is neither A26 nor S2, but another paper from the Guy's/Newcastle team, namely Pickup *et al*, 1979. This paper, which was slightly later than S2, was probably used as the key reference in the Kroc study because it described the use of pumps in a way that was nearer to the approach adopted in the Kroc study. This means that the Kroc study was not in our 2nd generation papers and, had we identified it through bibliometric methods, it could only have appeared in a later generation.

The importance of the Kroc study is that it is seen as a crucial step towards the National Institutes of Health (NIH) funded multi-centred Diabetes Control and Complications Trial (DCCT). The Kroc study concluded: 'These preliminary observations indicate the need for longer trials' (The Kroc Collaborative Study Group, 1984, p. 365). The Kroc paper, S9 in Alberti's list, is cited early in the main DCCT paper (DCCT Research Group, 1993). All interviewees agreed that the results from the DCCT have had a major impact on diabetes treatment in recent years by demonstrating that improved control over blood glucose and blood pressure leads to undoubted benefits in terms of reduced likelihood of complications. It is often difficult for databases accurately to record citations to papers whose authorship is a group. Nevertheless, this paper seems to have at least 870 citations recorded in various ways. There was wide agreement about the importance of the contribution of the Newcastle/Guy's collaboration to this stream of work.

In addition to informing a major stream of research, the Guy's/Newcastle work more directly fed into clinical practice in two ways. First, subcutaneous pumps have been refined as a way of delivering insulin and paper A26, according to one of the co-authors, might have influenced the commercial development of pumps because of its focus on bolus delivery at meal times. Pumps have been introduced more widely in the USA and the rest of Europe than in the UK, despite having been invented in the UK. The issue of pumps was recently investigated by the National Institute for Clinical Excellence (NICE). The Diabetes UK submission to NICE stated that less than 0.2% (or 650 people) with type 1 diabetes use pumps in the UK compared to about 8% in the USA, and 12% in Sweden. An article by Home *et al* (1982) is one of the papers cited in the submission as showing the benefits from pumps; this paper, in turn, cites three of the 1981 papers and so appears in our 2nd generation. Having itself been cited about 70 times it would have been an important 1st generation paper under our original plan of taking a three-year period.

In the recent guidance from NICE, insulin pump therapy 'is recommended as an option for people with type 1 diabetes provided that:

- Multiple-dose insulin (MDI) therapy (including, where appropriate, the use of insulin glargine) has failed; and
- Those receiving the treatment have the commitment and competence to use the therapy effectively' (NICE, 2003, para 1.1).

The guidance goes on to state that the view of the Appraisal Committee was that 'the proportion of people with type 1 diabetes who would be appropriate for, and would take up, insulin pump therapy, would be in the order of 1% to 2% of the total' (NICE, 2003 para 4.3.10). Nicely rounding off this discussion is the fact that the DCCT is one of the studies cited in the appraisal

consultation document from NICE as showing the benefits of maintaining good glycaemic control (NICE, 2002).

The second direct clinical impact linked to the stream of work on pumps is more controversial and is described in paper A8. The hypothesis tested in this Guy's/Newcastle paper was that for the diabetic patients with the most unstable glucose control, known as brittle diabetics, there was a physiological reason why the insulin was not being absorbed properly. The pumps were relevant in that they could be used to improve the administration of insulin. Brittle diabetic patients from around the country were taken on by Guy's and Newcastle. Tattersall described how he took a different view and always believed the causes of brittle diabetes were psychological not physiological. He suggested that this was now the accepted view. This was clearly a major debate in research relevant for clinical practice, and paper A8 has been fairly widely cited (44 times).

4.4.5 Studies of commercial relevance

Various studies were relevant for pharmaceutical companies and device manufacturers. They include: A7/S4; A22; A27; A13; A28; and A29. Despite the common commercial link, the nature of these studies varied.

Alberti described A7 and A22 on the Novo semi-synthetic human insulin as being 'very basic studies on pharmacokinetics'. There is wide agreement about the technical competence of these studies, and Murphy felt that some of the stream of such studies from Newcastle provided methodological developments that indicated how such studies should be undertaken in the future. Paper A27, in a similar area to that of A7 and A22, pioneered a new way of assessing the comparative efficacy of insulin preparations. However Tattersall, in particular, argued that the work in A7 and A22 was more in the nature of routine testing. Certainly in the critical pathway, Home notes that the large literature that followed was similarly stimulated by funding from the pharmaceutical companies. The quite high citation rates partly reflect the level of funding for further studies in these areas. The companies themselves were responsible for the clinical introduction of human insulin, which was so rapid that it was sometimes controversial. Nevertheless, studies such as these meant the companies could claim that the proper testing had been undertaken. (The studies did not show that semi-synthetic human insulin actually provided better results, but rather that there was no difference in clinical outcomes between using human and animal insulin).

In terms of research that flowed from these studies, Home describes how the studies gave an enormous push to the understanding of the problems and needs of subcutaneous insulin therapy

and, therefore, there is a clear link with the stream of work described in the preceding subsection. Papers A7 and A27 were quite highly cited—42 and 41 times respectively. Furthermore, these two, plus A22, all contained a below average number of citations where the importance of the cited paper to the citing papers was classified as Peripheral. This was particularly noticeable for A27 (where it was down to 19%) and the percentage of citations of high importance was the second highest, also at 19%—though four of the six were self-citations. The impact of this work, however, was perhaps rather concentrated because A7, A 22, and A27, also all had some of the shortest half-lives out of the 1981 set. A further point that might be of generic relevance but was claimed specifically in relation to these papers, is that two of the authors, Owens and Home, have been leading international and UK lecturers on insulin therapy for many years. Presumably, producing early, successful and reasonably highly cited papers in a field provides the credibility to be invited to lecture, and further citations might flow from this.

Paper A13, describing the testing of the slowing of carbohydrate absorption by acarbose, has the highest number of citations in the 1st generation. Furthermore, a below average number of these were classified as Peripheral. This partly reflects the fact that this was seen as a good basic study in the development of the drug, and one published in a journal with a high impact factor score. But it is also related to the history of acarbose. Shortly after this paper was published a study linked acarbose to cancer in rats, which stopped further development. Some years later this was found to be untrue; so it came back on stream and is now thought to have a potentially important role. Given this long history it is not surprising that paper A13 has one of the longest half-lives. One of the co-authors suggested that this paper was linked to the licensing of acarbose.

Paper A29 was one of a large number at this time testing different ways of conducting aspects of home blood-glucose monitoring. It is suggested it had some impact on the techniques developed by manufactures and used by practitioners. Paper S5 cites A29 and is therefore a 2nd generation paper. It demonstrates the important point that it was not so much the fact that it was blood glucose, rather than urine, monitoring that gave the advantage so much as the extra attention and information patients were receiving. Again, despite the agreement that this was an important finding, there was some disagreement about how the paper should best be interpreted in terms of clinical practice.

Paper A28 addressed the issue of meal-time insulin delivery in the context of testing its injection by a gun or jet. The reported findings about the limited advantages of using the gun, which was being heavily marketed, probably helped slow wider distribution: the findings were incorporated into guidelines. In the critical pathway it was argued that, along with paper A26, it was one of the

first to measure insulin profiles in plasma after injection, which has since become routine in such studies. Furthermore, it is suggested that the insulin profiles from A28 are still being used by some lecturers to illustrate the profile of human/pork insulin absorption after injection of meal-time insulin. These factors might explain why, despite the negative findings which are usually associated with short half-lives, in fact this paper had the third longest citation half-life of all 29 papers.

4.4.6 Studies that disproved an existing theory

Paper A23 describes one of the first of several studies that disproved the Chlorpropamide alcohol flushing hypothesis. As Home notes, this raises interesting issues in terms of payback because it was an important study that had an impact, but one that stopped a line of inquiry. The relevance of this interpretation is seen in the citation figures which are just above average, but which have the shortest half-life.

4.4.7 Other studies on the selected list and the overall link to clinical practice

Tattersall saw paper S7 as being important because it showed, probably for the first time, that raised blood glucose levels at the time of myocardial infarction indicate diabetes not transient “stress hyperglycemia”. By contrast, and illustrating well the point that much research represents small incremental steps, Tattersall suggested that paper S6 added only a little to something he had reported a couple of years earlier (Tattersall and Gale, 1981) as being part of his practice.

Roberts argued that the flow of studies from the Newcastle team had had a considerable influence on clinical practice, especially in the North East. She suggested that S6 was one example of that. She did not always think it appropriate to follow exactly the research findings from Newcastle and, as with the approach to surgery, sometimes developed her own method of implementing the approach. In this way there was more local ownership of the implementation. Nevertheless, the influence on clinical practice, which was of course national and international as well, was particularly strong in the Newcastle region. This was because of various factors including: the training of doctors who took up posts locally; the seminars etc run by the Newcastle team; and the role of Alberti as a respected opinion leader. Alberti himself acknowledged that there had been something of a ripple effect. There is some debate in general about the link between clinical and academic excellence, but both Alberti and Tattersall stressed its importance. Alberti was instrumental in securing resources for a Newcastle Diabetes Care Centre that concentrated the diabetes care in the city on a dedicated centre. In 1999 it won the Diabetes Centre of the Year Award. Examples of the international impact on clinical practice included ones from a co-author who referred to the influence on clinical guidelines from the Italian Society for the Study of Diabetes.

4.4.8 Impact on career development and medical education

Reference was made in Chapter One to the payback model that is informing this study. A key element in this is the research training that can be gained through participation in research projects (Buxton and Hanney, 1996; Buxton *et al*, 2000). Similarly, attention has recently focused on evaluating the social and economic benefits from publicly funded basic research by examining the mobility of scientists, particularly as they move into the commercial sector (Zellner, 2002). Home, in his critical pathway, was able to identify at least nine of the lead authors of the 1981 set of papers, including himself, for whom the involvement in this body of research provided a sound basis for subsequent career development. Similar points were made in interviews and by various co-authors in the questionnaires. One, for example, described the experience as, 'valuable training for the subsequent 110 papers/projects'. Though much of this came from the training received at Newcastle, there were also, as with paper A24, benefits from Alberti's wider links. Ralph Stevenson, the lead author of paper A24, is a notable example of a researcher who later moved into the commercial sector: he eventually became head of diabetes research at Pfizer.

Some of those from overseas who spent time in Newcastle also described the benefits to them in glowing terms. Generally, those who came from overseas were seen as good researchers, usually in the early stages of their careers. They often made a positive contribution to the portfolio of research at Newcastle.

The questionnaire to co-authors elicited a range of responses to the question about the contribution to medical training that was derived from the research. There was an overlap between discussion of the personal training received and the use of the findings in teaching. In addition to the points discussed above, several co-authors gave specific examples of where they thought the research had fed into medical education—again not just in the UK.

4.5 Key findings from the qualitative analysis

4.5.1 Key observations about the payback from the body of work

Attempting to describe the impact from the 1981 body of work, and that from the 10 selected papers, underlines the complex reality of how science advances and influences clinical practice. Many papers, if they make a contribution at all, make a small, incrementalist one. A few papers, however, have been shown to make a considerably greater impact. A possible key to the level of payback indicated is the enormous breadth of Alberti's contacts, and fields and methods of working. This is well illustrated in the account of how the idea for subcutaneous pumps came about. Similarly, perhaps, and as mentioned above, the ability to produce the very important guidelines on treating diabetics during surgery, and diabetic coma, partly resulted from the application to clinical problems of the understandings gained from some of the basic/early clinical

studies. What is significant in relation to the key papers S1, S2, and S3 is that they were having an impact on the work of 1981.

What perhaps is less clear from our study is an overall feel for how far the collection of papers from 1981 have been drawn upon in ways that would match these examples. Nevertheless, even here, paper A5, on treating diabetics during open heart surgery, and paper A26, on bolus delivery of insulin at meal times, are key parts of these wider streams. Furthermore, various papers, including A13 on acarbose, A24 on portal infusion of insulin, and A7 on semi-human insulin, have been shown to be important steps in bodies of work in their respective areas. Finally, the full complexity of the issues was summed up by paper A23 that helped to debunk the Chlorpropamide alcohol flushing hypothesis, and thus end a line of scientific enquiry.

4.5.2 Key lessons learnt about how to assess payback

On reflection, 1981 perhaps did not provide a really coherent starting point for this preliminary study despite it being relatively close to the time when Alberti moved to Newcastle. This was not only because of the large number of collaborations that he kept up, but also because important work that had been undertaken a few years earlier was still overlapping with the 1981 work.

However, the alternative approach of 10 selected papers, chosen from a longer period, is also problematic in various ways. For example, the omission of the paper cited in the Kroc study shows that not all the key links could come from analysis restricted to just the papers on such a list. Furthermore, the list of 10 papers also does not do justice to the full range of Alberti's work: it excludes, for example, a more basic *Lancet* paper (Alberti, 1973) that has had over 500 citations and much other important work, including studies related to Africa and more recent activities. (However, our study was never intended to review the full payback from the whole career of such a leading researcher). Looking at it from an opposite angle, however, such a list of selected papers might underplay the extent to which many studies have a very limited impact, if any.

Nevertheless, the variety of methods used has been successful in demonstrating a range of points. The relationship between qualitative analysis and bibliometrics is not simple, but there do seem to be benefits from combining the approaches, as was suggested by Gläser and Laudel (2001). The benefits come in various forms. The importance of the surgery papers despite their relatively low citation rates illustrates that analysis of citations alone is not sufficient, notwithstanding the fact that surgery papers tend on average to receive fewer citations than papers in the diabetes field (Wellcome Trust/NHS Executive, 2001). Furthermore, some of the more highly cited papers are revealed to be ones that are not necessarily included in those seen in the qualitative study as being of most significance. Some of the qualitative analysis, therefore,

shows bibliometrics is not sufficient. Other aspects of the qualitative approach, however, are strengthened by comparison with bibliometric data. For example, information on citation half-lives is providing useful correlations with some of the accounts from interviews etc. Similarly, comments in some of the interviews, and questionnaires, about the breadth of Alberti's collaborations correlate well with the discussion, in section 2.3.5, of the number of addresses per paper for the 1981 set being above average.

It is more difficult, however, to develop ways of fully utilising the extensive categorisation of the citations that was undertaken in Chapter Three. The figures for the importance of cited papers to the citing paper are difficult to link precisely to the qualitative analysis. There is some indication, however, that analysis of the importance of citations provides a better correlation with some of the comments made in the qualitative analysis than comes from crude citation counts alone. Evidence from the citation counting and categorisation is compatible with the view that a wide range of basic research is produced but only some of it eventually makes a significant contribution. The picture is, as yet, far from clear, but, nevertheless, some points are emerging. In any attempt to trace research through to its influence on practice, self-citations should certainly not be discounted in the way that happens in some assessments. Discounting self-citations would have meant, for example, not including the 1981 work that was cited in Home *et al* (1982) which, as noted, was eventually cited in the evidence from Diabetes UK to NICE on insulin pumps. While self-citations are not an illustration that peers regard the work as being worthy, the categorisation of citations suggests that self-citing can be an important way in which research findings are taken forward.

The contribution from the historical perspective provided by Tattersall also proved to be of great value, and this is consistent with those who advocate the use of an historical approach in examining the benefits from health research (Raiten and Berman, 1993; Berridge and Stanton, 1999). The significance of this contribution comes in several ways. On some issues, especially in relation to the treatment of brittle diabetics, the longer and wider perspective is useful in demonstrating that the implementation of some research findings is not always thought to be beneficial. But more generally, the historical approach helps to provide the realistically broad perspective within which the contribution of any one individual or team should be examined. It is interesting to note that while drawing on such a perspective, which often serves to reduce the significance of the contribution of an individual researcher, Tattersall was able to highlight various areas where Alberti's work had been of considerable importance.

5 LESSONS LEARNT AND THE WAY FORWARD

The three approaches used in this study have produced a large amount of data as reported in the previous chapters. Here we briefly review what we have learnt, describing the advantages and disadvantages to each approach, and make some overall assessments, before suggesting how the methods could be taken forward into a large-scale study as conceived at the outset of this preliminary project.

5.1 Lessons learnt about the exercise

We have used a combination of quantitative (bibliometric) and qualitative (case study) techniques to illustrate how a body of research published by George Alberti 20 years ago, in 1981, impacted on clinical practice in subsequent years. To do this we tested three methodologies:

- ***Citation tracing***

For the bibliometric analysis it was possible to trace the research through several generations, and then use these generations to conduct further bibliometric analysis including on the Research Levels of the generations, half-lives of the citations, and the number of authors and addresses on papers. It was possible to correlate some of this data with the qualitative analysis in ways that did illuminate various issues.

A major problem, however, was that, as with the previous attempts to work backwards, the number of papers rapidly became too large to handle. The original application had stated that it might be impractical to go beyond the 4th generation; as it turned out it was not possible to go beyond the 3rd generation. Had a similar pattern of growth occurred as happened between the 2nd and 3rd generations, the 4th generation would have consisted of at least 100,000 papers.

- ***Categorisation of citations***

The review of the literature on citation behaviour and categorising citations was utilised in the production of a template. The template was piloted by six assessors. The inter-rater reliability proved to be rather variable, but it was sufficiently satisfactory for one assessor to apply the template to the remaining citing articles. The ensuing categorisation provided some data that assisted with the interpretation of some comments made in the qualitative analysis. Furthermore, the extensive analysis has shed some light on the nature of citation behaviour. Self-citations to this body of work were, on average, categorised as being more important to the citing author than were citations given by other authors. Partly this arose from the nature of the extra authority that seems to accompany self-citations.

Additional work will be required to refine the template for citation categorisation, and the ways in which it is applied, before it is used more widely. Furthermore, given the time required to apply the template, plus the apparent lack of any overall pattern in terms of correlations between number and importance of citations, it would seem desirable to adopt a selective approach towards categorising citations.

- ***Qualitative study***

The qualitative analysis was able to produce a reasonably coherent account of the range of payback from the body of work examined. There was more impact on clinical practice than the study might have been expected to uncover, and considerable impact through research training was demonstrated. All the interviews produced useful material, and did so from a variety of perspectives, but the ability of an historian of diabetes to place Alberti's contributions within an historical analysis was of particular value. It is probably worth recommending that any future such analysis attempt to incorporate this perspective and the flexibility that it provides. That this important contribution came from one of Alberti's peers is also consistent with another key observation: having the insider expert from Newcastle to undertake the critical pathway analysis provided extremely valuable insights and detail.

It became clear in the qualitative phase that there were difficulties in undertaking critical pathway analyses, which restricted the number produced. Furthermore, there were so many 1st generation co-authors that it was impractical to interview them. Questionnaires, however, were successfully used instead.

- ***Overall assessment***

The variety of methods used has been successful in demonstrating a range of points. In some cases the additional methods are revealing things that a single approach could have missed. The importance of the surgery papers despite their relatively low citation rates illustrates that simple analysis of citations alone is not sufficient. Furthermore, some of the more highly cited papers are revealed to be ones that are not necessarily included in those seen in the qualitative study as being of most significance. Nevertheless, the use of multiple approaches is also helping to strengthen the analysis where points from one technique reinforce those from others. For example, some of the detailed bibliometric analysis, such as that on citation half-lives, is providing useful correlations with some of the other accounts from interviews etc. Martin and Irvine (1983) developed the concept of 'converging partial indicators' which involve the use of a range of indicators when considering the quality of teams of researchers operating in expensive areas of 'big

science'. Our combination of techniques suggests there are advantages in using a range of techniques when attempting to trace the impact from basic, or early clinical, biomedical research. Furthermore, the evidence from this study suggests it might be potentially useful to begin to look for correlations in relation to the breadth of work conducted by a research team, the levels of collaboration, the role of core funding support, and the degree of influence on clinical practice. So, for example, material from our interviews describes how Alberti believes in the benefits from undertaking a wide range of studies. The evidence from the analysis of Research Levels, and the qualitative analysis, demonstrates the breadth of Alberti's portfolio of research. Other bibliometric analysis highlights the large number of collaborators with whom he worked. And all this can be considered in the light of the analysis in Chapter Four indicating the impact made by this body of research.

What is proving more difficult, however, is to develop ways of fully utilising our extensive categorisation of the citations that resulted in the figures for the importance of the cited papers to the citing paper. On reflection, 1981 perhaps did not provide a really coherent starting point for this preliminary study despite it being relatively close to the time when Alberti moved to Newcastle. This was not only because of the large number of collaborations that he kept up, but also because important work that had been undertaken a few years earlier was still overlapping with the 1981 work.

Nevertheless, it is reasonable to conclude that whilst some difficulties were encountered with each approach, and some matters need addressing at an overall level, sufficient progress has been made to warrant considering how to take the study forward.

5.2 From preliminary to large-scale study

Various issues need to be addressed as we move towards a large-scale study. These include:

- ***Refining the techniques***

Further development and refinement of some of the methods will be undertaken as support is sought for the large-scale study. This will include attempting to improve the consistency of application of the categorisation of citations, and testing the ideas proposed below for applying the quantitative methods in a selective manner to key papers identified in the qualitative study. As a further preliminary step therefore, we will initially explore these activities in relation to the data already gathered in the study. Furthermore, we shall aim to incorporate improved techniques being developed elsewhere, including a new method of determining journal Research Levels (Lewison and Paraje, 2003).

- ***Defining the research questions***

As described in Chapter One, our study objective was to develop and test a number of methodologies. We therefore did not have a specific research hypothesis to test, beyond examining the assumption that basic research does feed through into eventual impact on clinical practice. However, in moving towards a large-scale study, we may want to focus on specific policy-relevant issues such as identifying and comparing factors that are associated with the successful translation of basic and early clinical research into clinical science and application. The preliminary study indicated several factors that could be included in the analysis. These relate to how far the production of research that is transferred into practice is connected to: the breadth of work undertaken; the number of collaborators; and support from core funding.

- ***Focus of the study***

In the current study we focused on one research group (Alberti) from one point in time (1981). If we expand the scope of the study (as suggested above) then it will be necessary to test the reliability and validity of any factors that are associated with, say, the successful translation of research. To give confidence in our findings, we will need to expand the focus of the study, to cover at least four sets of case studies. Ideally, each set of case studies should focus on the work of a number of research groups in the same field, in the same country. Our sets of case studies would cover two or three different fields, hopefully in at least two different countries.

5.3 Proposal for a large-scale study

If we increased the scope and focus of the study as suggested, some initial work would be necessary to select the case studies. For each case study we would have two methodological elements, qualitative and quantitative, which would work in parallel, each informing the other, to provide data for a research report.

Qualitative research

The qualitative aspect of a future large-scale study would be to develop an oral history of a research group's composition, key previous work of group members, the research programme in the early 1980s, and then trace the research originating from the group over a 20 year time period. To do this we propose to use a number of qualitative methods including:

- ***Group review***

We would need to inform ourselves of the work of the selected research groups by: reviewing all the literature published by the research groups, be that in the serial peer-reviewed literature or (where available) the 'grey' literature, contemporary commentators (for example editorials) and retrospective historical accounts (for example from the history of medicine literature); reviewing archival material from the research funders (such as original project proposals, referees' comments etc) and the researchers themselves; developing (with the help of 'field experts') critical pathways of the research group's outputs and contributions to career development.

- ***Key informant interviews***

We would need to try to interview the original members of the research group, their collaborators, and others who might know about the impact of the research. The interviews will cover the formation of the group and its research and, using the 'payback model', the long-term impacts of the research, including: contributions to clinical guidelines or other policies or health promotion schemes; uptake of the findings by industry; and influence of the research on medical training and clinical practice.

- ***Workshop***

For each case study we could run a workshop where we would invite a sample of the interviewees and other relevant stakeholders. The workshop could be run along the line of the Wellcome Trust's Witness Seminar Series to explore chronologically the development of research originating from the research group.

- ***Describe 'research lines'***

Using the primary and secondary resources from the document review, the key informant interviews and the workshop, it should be possible to piece together a detailed oral history of a case study's research programme and its impacts over a 20 year period. In addition, the identification of 'research lines' – that is discrete bodies or themes of research each addressing a particular question pursued by the research group – around the payback model would allow the application of standard comparable multidimensional indicators.

- ***Study seminar***

For all cases on the same topic there could also be a seminar at which the various time-lines would be discussed, in particular the overlaps between the research lines and

research groups. Whether such a seminar would cover the research in the field of more than one country would depend on the nature of the emerging findings.

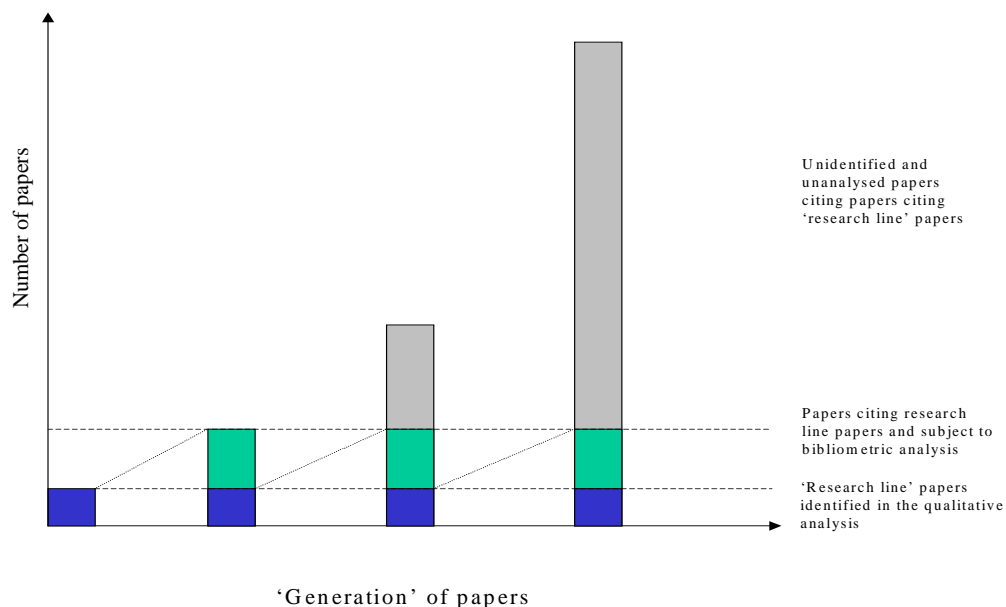
Quantitative bibliometric analysis

We would propose to use bibliometric data to inform and validate our qualitative research. The quantitative methods would include:

- ***Citation tracing for ‘research line’ papers***

Using the techniques reported in Chapter Two, we could develop a citation trace for ‘research lines’ identified from case study research. Initially this activity would start in conjunction with the ‘group review’ described above. To overcome the issue of exponential growth in our citing papers, however, we could – for each generation – focus our analysis on a core set of papers that were explicitly identified and referenced in the ‘research line’, and their ‘parental’ papers, as illustrated in Figure 16. By taking this type of approach we would be able to use the results of our qualitative study to focus our bibliometric analysis on the key citation trace through a research line.

Figure 16: Schema illustrating citation tracing for research line papers



- ***Categorisation of key papers***

We would need to refine the survey instrument for categorising papers, and test its inter-rater reliability. However, it could then be applied to the papers that cite the key papers from our research line.

Presenting the findings

Each research line could be written-up in a standardised document describing the impact of that research using the HERG payback model and categories (Buxton and Hanney, 1996). In addition, we would be able to describe a quantitative citation trace, using bibliometric techniques, for a research line and a randomly generated control. We shall use the qualitative and quantitative data to compare and contrast the 'payback' of research lines by country, disease and identify common factors that correlate with the translation of basic or early clinical research.

5.4 Concluding comments

The approach we have reported in this study is innovative and challenging. In the era of 'evidence based policy' research funders, as with all other sectors of society, are looking for value for money in the research they support. This means that they need to understand the effectiveness of different research strategies. This in itself poses a number of challenges, not least because the translation of new knowledge into improved policy and/or practice and commercial and non-commercial outcomes occurs over a long time period.

In this study we have begun developing a methodology that will allow us to understand the complexity of research development over a generation. The utility of the type of policy research proposed here will only be realised when it is scaled up to cover a number of different fields in different settings. This type of benchmarking study using comparative cases should allow the identification of factors that are common to, say, the successful translation of research. This in itself could inform research funders on their contemporaneous strategies, thereby increasing the likelihood of research getting into practice and thus increasing its effectiveness.

References

- Aknes DW (2003) A macro study of self-citation. *Scientometrics*, vol 56, 235-246
- Alberti KGMM, Christensen NJ, Christensen SE, Hansen AA, Iversen J, Lunbæk K, Sayer-Hansen K and Ørskov H (1973) Inhibition of insulin secretion by somatostatin. *Lancet*, 1299-1301
- Allen C, Qin J and Lancaster FW (1994) Persuasive communities: a longitudinal analysis of references in the *Philosophical Transactions* of the Royal Society, 1665-1990. *Social Studies in Science*, vol 24, 279-310
- Berridge V and Stanton J (1999) Science and policy: historical insights. *Soc Sci Med*, vol 9, 1133-1138
- Bonzi S and Snyder H (1991) Motivations for citation : a comparison of self citation and citation to others, *Scientometrics*, vol 21, 245-254
- Brooks T (1985) Private acts and public objects: An investigation of citer motivations. *Journal of the American Society for Information Science*, vol 36, 223-229
- Buxton M and Hanney S (1994) *Assessing payback from Department of Health Research & Development: Preliminary Report. Volume 1: The Main Report. Research Report No 19.* Uxbridge: HERG, Brunel University
- Buxton M and Hanney S (1996) How can payback from health services research be assessed? *Journal of Health Services Research and Policy*, vol 1, 35-43
- Buxton M and Hanney S (1997) *Assessing payback from Department of Health Research & Development: Second Report. Volume 1: Main Report. Research Report No 24.* Uxbridge: HERG, Brunel University
- Buxton M, Hanney S, Packwood T, Roberts S and Youll P (1999) *Assessing the benefits from North Thames Research and Development. Research Report No 25.* Uxbridge: HERG, Brunel University
- Buxton M, Croxson B and Hanney S (1999) *Assessing the Payback from Health R&D: From ad hoc Studies to Regular Monitoring. Research Report No 27.* Uxbridge: HERG, Brunel University
- Buxton M, Hanney S, Packwood T, Roberts S and Youll P (2000) Assessing benefits from Department of Health and National Health Service research and development. *Public Money and Management*, vol 20, 29-34
- Buxton M and Schneider W (1999) *Assessing the 'Payback' from AHFMR-funded research.* Alberta, Canada: Alberta Heritage Foundation for Medical Research
- Cano V (1989) Citation behavior: Classification, utility, and location. *Journal of the American Society for Information Science*, vol 40, 284-290
- Case D and Higgins G (2000) How can we investigate citation behaviour? A study of reasons for citing literature in communication. *Journal of the American Society for Information Science*, vol 51, 635-645
- Chubin DE and Moitra SD (1975) Content analysis of references: Adjunct or alternative to citation counting. *Social Studies of Science*, vol 5, 423-441

COHRED Working Group on Research to Action and Policy (2000) *Lessons in Research to Action and Policy – Case Studies from Seven Countries*. Geneva: The Council on Health Research for Development

Comroe JH and Dripps RD (1974) Ben Franklin and Open Heart Surgery. *Circulation Research*, vol 35, 661-669

Comroe JH and Dripps RD (1976) Scientific Basis for the Support of Biomedical Science. *Science*, vol 192, 105-111

Cozzens SE (1985) Comparing the sciences; citation context analysis of papers from neuropharmacology and the sociology of science. *Social Studies of Science*, vol 15, 127-153

Cozzens SE (1997) The knowledge pool: measurement challenges in evaluating fundamental research programs. *Evaluation and Program Planning*, vol 20, 77-89

Cronin B (1984) *The Citation Process. The Role and Significance of Citations in Scientific Communication*. Taylor Graham, London

Croxson B, Hanney S and Buxton M (2001) Routine Monitoring of performance: what makes health research and development different? *Journal of Health Services Research and Policy*, vol 6, 226-232

Diabetes and Control and Complications Trial Research Group (1993) The effects of intensive treatment of diabetes on the development and progression of long-term complications in insulin dependent diabetes mellitus. *New England Journal of Medicine*, vol 329, 977-987

Garfield E (1996) When to cite. *Library Quarterly*, vol 66, 449-458

Gilbert GN (1997) Referencing as Persuasion. *Social Studies of Science*, vol 7, 113-122

Gläser J and Laudel G (2001) Integrating scientometric indicators into sociological studies; methodical and methodological problems. *Scientometrics*, vol 52, 411-434

Grant J (1999) Evaluating the outcomes of biomedical research on health care. *Research Evaluation*, vol 8, 33-38

Grant J, Cotterell R, Cluzeau F and Fawcett G (2000) Evaluating “payback” on biomedical research from papers cited in clinical guidelines: applied bibliometric study. *British Medical Journal*, vol 320, 1107-1111

Grant J, Green L and Mason B (2003) *From Bedside to Bench: Comroe and Dripps Revisited. Research Report No 30*. Uxbridge: HERG, Brunel University

Hanney S, Davies A and Buxton M (1999) Assessing benefits from health research projects: can we use questionnaires instead of case studies? *Research Evaluation*, vol 8, 189-199

Hanney S, Gonzalez-Block M, Buxton M and Kogan M (2003) The utilisation of health research in policy-making: Concepts, examples and methods of assessment. *Health Research Policy and Systems*, vol 1(2) [<http://www.health-policy-systems.com/content/1/1/2>]

Home PD, Capaldo B, Burrin JM, Worth R and Alberti KGMM (1982) A crossover comparison of continuous subcutaneous insulin infusion (CSII) against multiple insulin injections in insulin-dependent diabetic subjects: improved control with CSII. *Diabetes Care*, vol 5, 446-47

- Jacob R and McGregor M (1997) Assessing the impact of health technology assessment. *Int J Technol Assess Health Care*, vol 13, 68-80
- Kostoff R (1998) The use and misuse of citation analysis in research evaluation, *Scientometrics*, vol 43, 27-43
- Latour B and Woolgar S (1979) *Laboratory life: the construction of scientific facts*. Beverly Hills, CA: Sage Publications
- Lewison G and Wilcox-Jay K (2003) Getting biomedical research into practice—the citations from UK clinical guidelines. Proceedings of the Ninth ISSI conference, Beijing, China, August 2003
- Lewison G and Paraje G (2003) The Classification of biomedical journals by research level. Proceedings of the Ninth ISSI conference, Beijing, China, August 2003
- Leydesdorff L (1998) Theories of citation. *Scientometrics*, vol 43, 5-25
- Mason B, Green L and Grant J (2001) Factors that lead to advances in neonatal intensive care: Comroe and Dripps revisited. *Research Evaluation*, vol 10, 121-127
- Maričić S, Spakventi J, Pavičić L and Pifat-Mrzljak G (1997) Citation context versus the frequency counts of citation histories. *Journal of the American Society for Information Science*, vol 49, pp530-540
- Martin B and Irvine J (1983) Assessing basic research: some partial indicators of scientific progress in radio astronomy. *Research Policy*, vol 12, 61-90
- McCain WK and Turner K (1989) Citation context analysis and aging patterns of journal articles in molecular genetics. *Scientometrics*, vol 17, 127-163
- Milbank Memorial Fund/The Cochrane Collaboration (2001) *Informing Judgement: Case Studies of Health Policy and Research In Six Countries*. New York: Milbank Memorial Fund/The Cochrane Collaboration
- Moravcsik MJ and Murugesan P (1975) Function and Quality of Citations. *Social Studies of Science*, vol 5, 86-92
- Moravcsik MJ and Murugesan P (1979) Citation patterns in scientific revolutions. *Scientometrics*, vol 1, 161-169
- Narin F, Pinski G and Gee HH (1976) Structure of the biomedical literature. *Journal of the American Society for Information Science*, vol 27, 25-45
- National Institute for Clinical Excellence (2002) *Appraisal Consultation Document: Continuous Subcutaneous insulin infusion for diabetes*. London: NICE
- National Institute for Clinical Excellence (2003) *Guidance on the use of Continuous Subcutaneous Insulin Infusion for Diabetes. Technology Appraisal Guidance-No.57*. London: NICE
- National Institutes of Health (2000) FY 2000 Research Program Outcomes Assessment Material. Bethesda, MY: NIH
- Oppenheim C and Renn SP (1978) Highly cited old papers and the reasons why they continue to be cited. *Journal of the American Society for Information Science*, vol 29, 225-231

OST (1993) (White Paper) Realising our Potential – A Strategy for Science, Engineering and Technology. Command 2250. London: HMSO

Pang T, Sadana R, Hanney S, Bhutta Z, Hyder A and Simon J (forthcoming) Knowledge for better health—A conceptual framework and foundation for health research systems. (Submitted to the *Bulletin of the World Health Organization*)

Pemberton MJ and Nugent CR (1995) Information studies. Emergent field, convergent curriculum. *Journal of Education for Library and Information Science*, vol 36, 126-138

Phoolcharoen W (2002) *Health Systems Reform in Thailand: The Role of the Health Systems Research Institute. Alliance-HPSR Working Paper No. 14*. Geneva: Alliance for Health Policy and Systems Research

Pickup JC, White MC, Keen H, Parsons JA and Alberti KGMM (1979) Long-term continuous subcutaneous infusion of insulin infusion in diabetics at home. *Lancet*, vol 2, 870-873

Pichappan P and Saravady S (2002) The other side of the coin: the intricacies of author self-citations. *Scientometrics*, vol 54, 285-290

Prabha C (1983) Some aspects of citation behavior: A pilot study in business administration. *Journal of the American Society for Information Science*, vol 34, 202-206

Raiten D and Berman S (1993) *Can the Impact of Basic Biomedical Research be Measured?: A Case Study Approach*. Bethesda MD: Life Sciences Research Office, Federation of American Societies for Experimental Biology

Royal Netherlands Academy of Arts and Sciences (2002) *The Societal Impact of Applied Research Towards a Quality Assessment System*. Amsterdam: Royal Netherlands Academy of Arts and Sciences.

Shadish WR, Tolliver D, Gray M and Sen Gupta SK (1995) Author judgments about work they cite: Three studies from psychology journals. *Social Studies of Science*, vol 25, 477-498

Small H (1978) Cited documents as concept symbols. *Social Studies of Science*, vol 8, 327-340

Small H (1982) Citation context and content analysis. In Dervin B and Voight M (eds) *Progress in Communication Sciences*. Norwood, N.J: Ablex, vol 3, 287-310

Smith R (2001) Measuring the social impact of research. *BMJ*, vol 323, 528

Sönksen PH, Srivastava MC, Tompkins CV and Nabarro JDN (1972) *Lancet*, vol ii, 155

STATA Press (1999) *STATA Reference Manual: Release 7.0*. College Station, TX: STATA Press

Spiegel-Rosing I (1977) *Science Studies: Bibliometric and Content Analysis*. *Social Studies of Science*, vol 7, 97-113

Tattersall RB and Gale EAM (1981) Patient self-monitoring of blood glucose and refinements of conventional insulin treatment. *American Journal of Medicine*, vol 70, 177-182

van Weel C (2002) Biomedical science matters for people—so its impact should be better assessed. *Lancet*, vol 360, 1034-1035

Vinkler P (1987) A quasi-quantitative citation model. *Scientometrics*, vol 12, 47-72

Vinkler P (1998) Comparative investigation of frequency and strength of motives towards referencing. The reference threshold model. *Scientometrics*, vol 43, 107-127

Von Ungern-Sternberg S (2000) Scientific communications and bibliometrics, Abo Akademi University, Finland. [www.abo.fi/~sungern/comm00.htm – accessed on 17/05/01]

Wellcome Trust (1998) *Mapping the Landscape: National Biomedical Research Outputs 1988-1995*. London: The Wellcome Trust

Wellcome Trust/NHS Executive (2001) *Putting NHS Research on the Map: An Analysis of Scientific Publications in England 1990-1997*. London: The Wellcome Trust

White M and Wang P (1997) A qualitative study of citing behavior: Contributions, criteria, and metalevel documentation concerns. *Library Quarterly*, vol 67, 122-154

World Health Organization (2002) *National Health Research Systems – Report of an International Workshop*. Geneva: World Health Organization

Zellner C (2002) Evaluating the social economic benefits of publicly funded basic research via scientists' career mobility. *Research Evaluation*, vol 11, 27-35

Appendix 1: Funding sources for each paper

Paper Reference Number	Sources of Funding
A1	None
A2	Synthelabo (Paris); British Diabetic Association
A3	The Wellcome Trust; British Diabetic Association
A4	British Diabetic Association; Newcastle Area Health Authority
A5	None
A6	British Diabetic Association; Medical Research Council (UK)
A7	Novo Industries
A8	British Insulin Manufacturers; Medical Research Council (UK); Novo Ltd; Minet Trust; British Diabetic Association
A9	Newcastle Area Health Authority; British Diabetic Association; ; Medical Research Council (UK); The Wellcome Trust
A10	British Diabetic Association; The Wellcome Trust
A11	British Diabetic Association
A12	None
A13	Bayer UK; British Diabetic Association; Medical Research Council (UK); The Wellcome Trust
A14	Novo Industries; British Diabetic Association
A15	British Diabetic Association; Newcastle Area Health Authority (Teaching)
A16	Royal Victoria Hospital Belfast Research Fellowship; British Diabetic Association
A17	British Diabetic Association; The British Council; Novo Industries
A18	None
A19	British Diabetic Association
A20	British Diabetic Association; The Wellcome Foundation (Trust?)
A21	Danish Medical Research Council
A22	Novo Industries
A23	[Pfizer Ltd materials]; Northern Regional Health Authority; British Diabetic Association
A24	British Diabetic Association
A25	British Diabetic Association; The British Council; Novo Industries, Denmark
A26	Servier Laboratories; National Institutes for Health; Novo Laboratories; British Insulin Manufacturers; British Diabetic Association
A27	Eli Lilly
A28	Novo Laboratories Ltd
A29	British Diabetic Association; Newcastle Area Health Authority (Teaching); [Medistron Ltd and Boehringer Corporation Ltd for the modified Glucocheck meter]

Appendix 2: Guidance notes for completion of the datasheet

We are assessing the “payback” from research as part of an NHS Executive-funded project and are collecting information on the relationships between cited and citing articles. You have been provided with one datasheet per citing reference and a copy of the citing article. The following notes are provided for your guidance but if you need further clarification on completing this datasheet please contact Iain Frame on 020 7611 8235 or e-mail i.frame@wellcome.ac.uk

- The cited article details are given at the top of the datasheet. Locate the cited reference details in the references section at the end of the article and mark it. Depending on the journal, references are cited in different ways. Some journals cite articles by all authors, some by first author (or first three authors) followed by *et al*, and some use a numbering system where the articles cited are numbered either in the order they appear or alphabetically by author.
- You should be aware that prior to 1985, the ISI Science Citation Index gives author details in the following order- first author followed by all other authors in alphabetical order. You may therefore find that the details on the datasheet don't match up exactly with the details in the references section of the paper.
- Locate the citations by reading through the article, marking where each citation occurs. As far as possible, try to fit the location into one of the categories provided (eg Introduction, Materials and Methods, Results or Discussion)
- When recording individual citation details, you should mark each individual citation, even when it occurs twice in the same sentence or paragraph.
- For each citation occasion you may tick more than one box according to the definitions provided.
- When assessing the overall importance of the cited article to the citing article, please use the following definitions for guidance. In cases where there is more than one datasheet per citing article, please answer the additional question relating to the overall importance of the body of work as a whole according to the definitions given below.

PERIPHERAL: The work described in the cited article is of little importance to the citing article. Citation is simply background, an aside, for completeness or indeed irrelevant.

LIMITED: The work described in the cited article is of some limited importance to the citing article. It would be inappropriate to omit it, but it is not an important part of a central argument.

CONSIDERABLE: The work described in the article is of considerable importance to the citing article. The work is one of a number central to the argument.

ESSENTIAL: The work described in the cited article is of critical importance to the citing article, and central to the argument presented, and a key foundation for the paper.

Appendix 3: Template for categorising articles

Cited Paper ID: A6

Authors: Barnett-AH Alberti-KGMM Burrin-J Pyke-DA Spiliopoulos-AJ Stubbs-WA

Title: Metabolic Studies in Unaffected Co-Twins of Non-Insulin-Dependent Diabetics

Journal: BRITISH MEDICAL JOURNAL **Vol:** Vol 282 **Pages:** pp 1656-1658

Addresses: UNIV LONDON KINGS COLL HOSP, DEPT DIABET, LONDON SE5 9RS, ENGLAND/ST BARTHOLOMEWS HOSP, DEPT MED, LONDON EC1A 7BE, ENGLAND/ROYAL VICTORIA INFIRM, DEPT CLIN BIOCHEM & METAB MED, NEWCASTLE TYNE NE1 4LP, TYNE & WEAR, ENGLAND

Citing Paper ID: B1

Title: Metabolic Studies in Unaffected Co-Twins of Non-Insulin-Dependent Diabetics

Authors: Hockaday-TDR

Journal: BRITISH MEDICAL JOURNAL

Individual citation details

For each occasion the article is cited, note its location (eg Introduction, Materials and Methods, Results or Discussion) in the citing article.

For each citation, tick the appropriate box(es) using the following definitions:

Develop: The *citing* article is developing a concept or method previously described in the cited article.

Support: The *citing* article is supporting a concept or method previously described in the cited article.

Apply: The *citing* article uses a method (or methods) described in the cited article.

Refute: The *citing* article either claims that the cited article is incorrect or disputes the cited article but is unable to come to a firm conclusion.

Note/Review serves The *citing* article refers to the cited article as part of the relevant literature but it either serves

Only: no explicit role in the analysis (note) or is compared to other relevant literature (review).

Citation occasion	Location	Develop	Support	Apply	Refute	Note/Review only
1 st		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2 nd		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3 rd		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4 th		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5 th		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6 th		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Additional citations ...

Articles class: Basic Clinical

Does the article describe **basic research** (ie theoretical or laboratory studies) or **clinical research** (ie relating to the observation and/or treatment of human subjects with a view to improving human health).

What is the overall importance of the cited article to the citing article?: Peripheral Limited Considerable Essential
Tick **one** box which you think is the most appropriate.

Appendix 4: Template for categorising reviews

Cited Paper ID: A28
Authors: Taylor-R Alberti-KGMM Home-PD
Title: Plasma Free Insulin Profiles After Administration of Insulin by Jet and Conventional Syringe
Journal: DIABETES CARE **Vol:** Vol 4 **Pages:** pp 377- 379
Addresses:

Citing Paper ID: B1020
Title: The Absorption of Subcutaneously Injected Insulin
Authors: Demeijer-PHEM Lutterman-JA Vantlaar-A
Journal: NETHERLANDS JOURNAL OF MEDICINE

1.3 Individual citation details

For each occasion the article is cited, make a judgement whether the citation lies in the introduction, conclusion or elsewhere in the article.

For each citation, tick the appropriate box(es) using the following definitions:

Support: The *citing* article is supporting a concept or method previously described in the cited article.

Refute: The *citing* article either claims that the cited article is incorrect or disputes the cited article but is unable to come to a firm conclusion.

Note/Review Only: The *citing* article refers to the cited article as part of the relevant literature but it either serves no explicit role in the analysis (note) or is compared to other relevant literature (review).

Citation occasion	Location	Support	Refute	Note/ Review only
1 st		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2 nd		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3 rd		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4 th		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
5 th		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6 th		<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Additional citations ...

What is the overall importance of the cited article to the citing article?: Peripheral Limited Considerable Essential
 Tick **one** box which you think is the most appropriate.

Appendix 5: Location and category of the citation occasions in five papers in the Essential group

Cited paper ie 1 st generation	Citing paper ie 2 nd generation	No of citation occasions	Location of citation				
			Introduction	Materials & Methods	Results	Discussion	Elsewhere*
A20	B152	9	4 Note/Review	-	-	3 Support 2 Refute	N/A
A20	B157	6	1 Note/Review	-	-	5 Support	N/A
A23	B65	6	1 Note/Review 2 Refute	N/A	N/A	N/A	1 Note/Review 2 Refute
A24	B175	8	1 Note/Review	2 Support and Apply**	-	5 Support	N/A
A8	B31	7	2 Support	2 Apply	1 Support	2 Support	N/A

* Location classification for review articles

** 2 citations judged to be both Support and Apply

Appendix 6: Questionnaire to original authors

1. What is your impression of the ways, if any, in which the paper(s) you co-authored with Sir George Alberti influenced subsequent scientific research by yourself or others?

2. Are you aware of any papers that gave a citation, or self-citation, to your paper(s) and that turned out to be particularly significant in terms of the development of a scientific line of enquiry?

Yes <

No <

If so, please give the reference of the citing paper if possible:

3. Are you aware of any impact that your paper(s) has had directly, or through further research by yourself and/or others, on eventual policy and/or practice, for example contributing to: national or local policies including clinical guidelines and health promotion schemes; or changes in the behaviour of practitioners at the national or local level?

Yes <

No <

If yes, please give examples:

4. Are you aware of any patents, or other commercial products, to which the research described in your paper(s), or in papers citing it, contributed?

Yes <

No <

If yes, please give details:

5. Are you aware of any contribution to medical training that was derived from the research described in your paper(s), or in research building on it?

Yes <

No <

If yes, please give details:

6. Apart from those appearing on the enclosed Table of 1981 co-authors, are you aware of any senior overseas researchers attracted to undertake a period of research in the UK as a result of the stream of work at Newcastle conducted by Sir George and his team?

Yes <

No <

If so, please give details:

7. Are there any other observations you would care to make about the overall body of work contained in the set of 1981 papers in the Table?

Appendix 7: Critical pathway instructions

Professor George Alberti's scientific output in 1981 included the attached 29 articles. We would like you to construct a critical pathway exploring the impact of this work in terms of the subsequent scientific advances and the impact on **health**.

It might be appropriate to start with a rapid review of the papers to decide:

- a) which papers have had the greatest impact on health; and
- b) is it possible to group the papers into perhaps 4 or 5 major areas (possibly identifying a key paper to represent each group) and continue the main analysis on that basis.

If a number of pieces of published work are related it would also be useful to provide a terminology (e.g. '*methods of insulin delivery*') for the Wellcome Trust/HERG research team to use to describe these groups of papers.

As guidance to undertaking the main analysis it may be useful to think first about the impact the scientific knowledge represented in these papers has made on subsequent scientific advances, whether made by the original authors or by other teams. Then it would be helpful to consider whether these paper (or groups of papers) have had impact on any of the following: national or local policies including clinical guidelines/health promotion schemes; the medical education provided at Newcastle or elsewhere; local or national clinical practice; commercial exploitation etc.

We have deliberately not made these instructions too prescriptive, it would therefore be helpful if you could briefly detail the methodology you used to perform the task.