MJM

AN INTERNATIONAL FORUM FOR THE ADVANCEMENT OF MEDICAL SCIENCE BY STUDENTS

MJM

AN INTERNATIONAL FORUM FOR THE ADVANCEMENT OF MEDICAL SCIENCE BY STUDENTS

VOLUME 7 • NUMBER 2 • 2004

EDITORIAL BOARD

MJM

AN INTERNATIONAL FORUM FOR THE ADVANCEMENT OF MEDICAL SCIENCE BY STUDENTS

ACKNOWLEDGMENTS

The following organizations, companies, and institutions deserve recognition for their generous contributions to the *McGill Journal of Medicine*.

McGill University Faculty of Medicine

McGill University Medical Students Society

Kenmara Incorporated

Type and Graphics Incorporated

Pfizer Canada Incorporated

The MJM wishes to sincerely thank Dr. Phil Gold, our faculty advisor, and Dr. Abraham Fuks, Dean of McGill Medicine, for their guidance, support, and continuing effort on behalf of MJM. We also wish to thank Drs. Moishe Liberman, David Mulder, and John Sampalis for contributing to this issue's Focus on Trauma.

The *MJM* is technically supported and powered by MedIT.

MJM imaging and printing by Type and Graphics Inc.
MJM hardware provided by the generous donation of McGill Faculty of Medicine
MJM advertising by Kenmara Inc.

The McGill Journal of Medicine (MJM), published by students at McGill University (Montreal, Canada), is devoted to the advancement of medical science by students. Materials in the MJM published by students include Original Articles, Case Reports, Review Articles, and Crossroads. Articles in the MJM Focus are the invited contributions of physicians and scientists. Letters to the MJM, Commentaries, and Book Reviews are submitted by all who wish to contribute. All materials reflect the individual views of the authors and do not represent the official views of the MJM or the institutions with which the authors are affiliated. Articles published in the MJM have been reviewed for accuracy; however, information presented and opinions expressed remain the sole responsibility of the author(s). Advertisements appearing herein do not reflect endorsement by the MJM. Although all advertisements have been reviewed to ensure compliance with general ethical and legal standards, ultimate responsibility for their content rests solely with the advertiser. The physician is therefore advised to verify all prescribing information. The Journal is published biannually (Summer and Winter) in one volume (two issues) per year. Student subscription prices for each volume are CAN$20(Canada), US$15 (USA), or US$20 (International). Non-student subscription prices are CAN$40 (Canada), US$30 (USA), or US$40 (International). Communications regarding advertising, subscriptions, and other business matters should be directed to the MJM Secretarial Division (secretarymjm.med@mcgill.ca), and inquiries regarding manuscript submissions and general editorial matters should be directed to the Editorial Division, at mjm.med@mcgill.ca. The MJM head office is located at the Faculty of Medicine, 6th Floor, 3655 Promenade Sir William Osler, Montreal, QC, Canada H3G 1Y6. The online version of the MJM is available at: http://www.mjm.mcgill.ca.

ISSN 1201-026X

MJM

Pfizer Canada, Inc.

Scientific Publication Award

First Prize $1000

Second Prize $500

CONTEST RULES:

- Prizes will be awarded to STUDENT AUTHORS of the top original research articles submitted for publication in the MJM issues 8.1 and 8.2
- Further details about the contest are available at http://www.medicine.mcgill.ca/mjm/pfizer/pfizeraward.htm
- Deadline for submission to MJM 8.1: June 1st, 2004
- Deadline for submission to MJM 8.2: January 15th, 2005

McGill Journal of Medicine
Faculty of Medicine, McGill University
3655 Promenade Sir William Osler
Montreal, Quebec
H3G 1Y1 CANADA
E-mail: *mjm.med@mcgill.ca*

Pfizer

www.mjm.mcgill.ca

McGill 2004
Student Bioethics
Conference

October 1st and 2nd 2004 - Montreal, Quebec, Canada

While genetic information and its associated technologies have had a major impact upon society, few of us have had the opportunity to discuss its future impact with leading experts.

Take part in expert discussions on

- *Cloning and stem cell technology*
- *Property rights on genetic information*
- *Patent rights on individual genes*
- *Genetic counseling and therapy*
- *International laws and policies*

Speakers

*Dr. Ruth Fischbach**
Dr. Kathleen Glass
Dr. Jonathan Kimmelman
Dr. Bartha Knoppers
Dr. Tom Hudson
Dr. Glenn McGee and others*
**to be confirmed*

For more information: http://bioethics.sus.mcgill.ca/
Email: bioethiconference@sus.mcgill.ca

EDITORIAL

HISTORICAL FOOTPRINTS: MJM PAST AND PRESENT

There has been a long tradition in the history of medicine to cement one's legacy by naming some anatomic structure, disease process, or new treatment after oneself; although to be fair, sometimes the naming was done posthumously by a third party. Either way, the net result has been a one-to-one association between researcher and discovery that has become ingrained into our collective consciousness. Of course, if things were that simplistic then the sheer number of medical eponyms that bear Virchow's name would probably lead us to believe that he was the most industrious man in the history of western medicine. The reality though, is that every great scientific discovery is the product of entire teams of researchers that rarely get their names mentioned in textbooks. So it would be unfair of us to talk of the Calvin cycle and credit its discovery to Melvin Calvin without at least giving a passing nod to Andrew Benson who did a considerable portion of the work while on Calvin's team. Also, while Denis Burkitt described the new form of lymphoma that would come to bear his name, it was Michael Epstein, Yvonne Barr, and Burt Achong that actually isolated the causative virus. Unfortunately, the naming of the Epstein-Barr virus pays little tribute to the third member of this triad.

This tirade against the injustices of historical nomenclature has its origins in the unlikely developments stemming from the Faculty of Medicine's annual newsletter. The newsletter carried an article about the MJM and its growing readership that provoked a rather surprising response. Two of our noted alumni, Dr. Charlotte Ferencz and Dr. William Gibson, wrote to us to point out that the article did little to recognise the previous incarnation of our medical journal. It is an accusation we readily acknowledge. In fact, our predecessor publication bore the slightly different name "McGill Medical Journal" and was first published in 1931. Its origins were far older than we had originally realised and its longevity was just as surprising. Its final issue, published in 1981 marked the end of its half-century run. However, the discoveries did not end there. It was three hard-working students, from the illustrious Class of '32, that originally founded the McGill Medical Journal. They were Clement Clay, James Gray and Colin McLeod. It was the last name that sparked the greatest amount of interest.

McLeod graduated from McGill in 1932 and spent the next two years as a resident at the Montreal General Hospital, before moving to New York and accepting a position at the Rockefeller Institute Hospital. It was then that he began to work with his fellow Canadian-in-exile Oswald Avery. The result of that partnership was the groundbreaking 1944 paper "Studies of the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types. Induction of Transformation by a Deoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III" by Avery, McLeod, and McCarty. The paper challenged the belief that genetic material was carried by proteins and proposed that it was in fact transmitted by these so-called "nucleic acids." Though McLeod did much of the actual physical work, he still had to play second fiddle to the more senior Avery who was director of the laboratory. However, the point became mute after Watson and Crick published their 1953 paper on the double helix (which incidentally never cited the work of Avery and McLeod). The magnitude of the publicity surrounding the double helix eclipsed the work done at the Rockefeller Institute, forever denying Avery and MacLeod a much-deserved Nobel Prize. In the 50th anniversary edition of Nature (vol. 421, no. 6921, p. 406, 23 January 2003), an editorial note called the omission "an oversight that, to this day, still puzzles."

So what is to be done with all these historical oversights? Perhaps, very little can be done. A drastic re-write of medical history would be cumbersome and serve no real purpose. All we can do is continue with the basic principles that have guided scientific researchers for the greater part of modern history. We provide a forum for scientific discussion and let everyone submit their papers, hypotheses, theories, and suspicions. There will be the inevitable amount of retractions, confusion, disputes, and competing claims over some work. We can only hope history will sort it out and give us some final pronouncement of who, if anyone, was ultimately correct. Morever, in this tempest of scientific publication, and in a world where big discoveries increasingly seem to require big research budgets, student research can often get lost in the whirlwind. Fortunately, there is a place where student research is always welcome. It is a place where great careers can begin, and it is place where researchers that will change the course of history can get their feet wet. That place is here at the MJM, and we are very proud of it.

REFERENCES

1. McCarty, M. Discovering genes are made of DNA. Nature 2003; 421: 406.

Kathy Han, B.Sc. (Hon.), M.D.C.M. (2006) and **Christopher Labos**, M.D. C.M. (2006) are the eighth Editors-in-Chiefs of the MJM. Kathy Han completed her B.Sc. in Immunology at the University of Toronto. For the past four years, she has been actively involved in cancer research, which has led to publications in *Nature Medicine* and *Molecular Cancer Research*. Christopher Labos has been awarded the Shirley Nancy Edman Prize for Physiology and earned the Ivan Racheff Scholarship for his research in the surgical treatment of cardiac tamponade.

LETTERS TO THE MJM

MEDICINE IN THE 21ST CENTURY

Medical students are no longer taking a passive role in their education. Medicine in the 21st Century*, a publication by United Kingdom (UK) Medical students, attempts to open the way for greater student involvement in educational reform and reflects a growing international trend.*

In 1993, the General Medical Council (the governing and licensing body for UK doctors) published clear guidelines on the required outcomes that medical students should meet upon graduation (1). These principles acknowledged the need to tailor undergraduate education towards developing a learning process and the skills of interaction with patients and colleagues, thereby building foundations for lifelong learning and effective patient-physician partnerships. Such broadening of the doctrine led to increasing diversity in undergraduate medical curricula with an emphasis on student-selected components. This provided the opportunity for students to govern some of their education, while simultaneously guaranteeing that the core competencies required of the profession would also be met. Such a radical revision of the foundations of the profession provoked further innovative thinking outside of the mainstream infrastructures including serious considerations of the potential merits of a virtual medical school (2).

In addition to these changes, increasing numbers of UK medical schools are updating their courses. Some are moving towards problem based and skills orientated teaching, where the emphasis is on the integration of core knowledge with real life patient scenarios, and encouraging the development of thought processes particular to the task of being a doctor. Entrance to the profession and the arena in which learning is achieved is also shifting with the advent of single entry foundation courses for health professionals, and the devolvement of some teaching responsibilities from hospitals to primary care.

But what about the 'consumers' of medical education? Are future doctors simply passive, detached observers of these processes of reform, or are they using the opportunities of change to input their own views? North America in particular has seen dynamic activity on this front. The American Medical Students Association (AMSA) has it's own “bill of rights” in which it states: “Medical students have the right to shape the content of their education, they practice what they preach”. In this case, the various official bodies and working groups involved in ongoing curriculum development are augmented by a strong student presence at both local and national levels so as to influence and hopefully improve the learning experience for their colleagues. AMSA has instituted many successful reforms, for example, the recent acquisition of a grant from the National Institutes of Health (NIH) to start teaching complementary and alternative medicine in medical schools.

Here in the UK, we have not been idle either; we have seized the opportunities that the aforementioned reforms have granted us. In the UK, medical students have made sure that their voice is heard on a whole range of issues. The British Medical Association (BMA), which is both the UK doctor's trade union and a professional body, has its own Medical Students Committee (MSC), which has over 13,000 members, and represents the UK's biggest single voice for future doctors. Past successes have included the addition of students to the official quality inspecting teams which visit and inspect medical schools, improving equity of treatment of students with health difficulties, and the contribution by students to key negotiations of the content and structure of undergraduate curricula.

The BMA's MSC recognised the need to reflect and input on the implementation of official recommendations set out by the GMC. This led to the genesis of "our own" policy document titled Medicine in the 21st Century. In this publication, the MSC sets out students' rights and responsibilities as professionals in training, delving further than existing guidance on standards in medical education. This is achieved by being specific in areas where the official guidance is too broad, and by looking at education holistically as part of a wider experience of living and training as future professionals. The spectrum of subjects covered includes: funding, careers advice, pastoral care, occupational health advice, and research. Medicine in the 21st Century was sent out for consultation last year and received positive feedback from the key players in UK medical education, including the Council of Heads of Medical Schools (the authoritative voice of UK Medical school leaders) and the General Medical Council. However, some were concerned that the document placed a disproportionate emphasis on students' rights, and said less about their responsibilities. The theme of the document remains to provide a simple but realistic agenda for change; one of partnership between current professionals and future professionals so as to provide for the basic academic and non-academic needs of medical students.

It is envisaged that Medicine in the 21st Century will be more than a dry policy statement, but an effective lever for change at both a national and local level.

Indeed, it is already opening discussions with the regulators of medical education. Students currently involved in feedback and assessment of existing courses will be best placed to use the document to suggest improvements to their Deans, curriculum design teams, staff-student committees, and similar bodies. Furthermore, the publication of the document on the BMA medical students website, will allow more medical students to access it, and highlight any deficiencies in their own training programmes, encouraging them to raise these matters locally.

The broad scope of Medicine in the 21st Century sets it apart from other documents on standards in medical education. Not only does it cover the "bread and butter" subjects of curriculum content, finance, welfare, and admission policies, but also several current "hot topics" including pre-graduation careers advice, elective research, and postgraduate study. Students have a unique perspective on education. It is sometimes difficult for those designing curricula and writing legislation, to appreciate the end product, its interpretation, and its shortfalls at grass-roots level. This perspective needs a comprehensive and coherent voice; this is the heart of Medicine in the 21st Century. The emphasis on students' responsibilities, as well as their rights, flows throughout. It states: “The stresses and strains of the medical degree can be considerable; learning how and where to seek help is an important skill that will benefit students not only as an undergraduate, but also throughout their medical career.” This emphasises the importance of professionalism and self care.

The issues surrounding equal opportunities for students in medicine has been a source of contention in the past, particularly regarding physical disability (3), Medicine in the 21st century explicitly calls for, “Medical schools [to] ensure that appropriate measures are in place to enable students with disabilities, including those suffering from dyslexia, to complete the course…medical schools should show flexibility and innovation when determining the process by which curriculum outcomes are met.”

The issue of equal opportunities doesn't end there. The document is the first of its kind to finally acknowledge the religious and cultural diversity of those studying medicine, and the implications this may have. Medicine in the 21st Century states; “Medical students should not be penalised for participation in religious or cultural events…[but] should give prior notice of their religious commitments when they impinge on the medical course”.

The constant evolution of medical education reflects a need for the relevance, effectiveness, and values of medical education to keep up with scientific advance, as well as the expectations of society; this way, tomorrow's doctors will be equipped to function safely and effectively in tomorrow's world. Medical students as professionals in training, stakeholders in the education they receive, and as grass-roots consumers of medical education, must seize this opportunity to argue for full participation in the development and implementation of new initiatives. In Medicine in the 21st Century, UK medical students have set out their own ideas of how to move forward in improving the overall package of education. It is imperative that students continue to be encouraged by senior medical educators in engaging with these issues, and contributing their unique perspective.

Sincerely,

Nicholas France, Bruno Rushforth and Rameen Shakur
The University of Warwick Medical School, Manchester University, and the Universty of Edinburgh respectively.

Competing interests: BR and RS were on the BMA MSC medicine in the 21st Century Working group. NF was deputy chair of the MSC 2002-2003. BR and NF have served as observers on the GMC's Education Committee

REFERENCES

1. General Medical Council. Tomorrow's doctors. London: GMC, 2002.
2. Barratt H. Virtual medical school will soon be a step closer to reality. student BMJ 2002;10:309. (September).
3. McDonald R, Discrimination in Medicine. BMJ 22 June 2002; 324:S199

Nicholas France holds an honours degree in Pharmacology from Kings College University of London and is currently a final year medical student at The University of Warwick Medical School. He is the previous deputy chairman of the BMA (British Medical Association) Medical Students Committee, current deputy chair of the Medical Students Conference and student observer to the General Medical Council (GMC) Education Committee. **Bruno Rushforth** holds a B.A (Hons) in Philosophy from Cambridge University and an M.A. in Health Care Ethics and Law from Manchester University where he is currently in the final year of his medical studies. He is employed by the GMC as an Education Visitor and has also been a student representative on the BMA Medical Students Committee where he chaired the working group that drafted *Medicine in the 21st Century* in 2001-2002. **Rameen Shakur** is a Medical student at the University of Edinburgh; he holds a BA and an MPhil from Cambridge University and is a Churchill Fellow at Harvard University.

LETTERS TO THE MJM

12 TIPS FOR POTENTIAL MJM AUTHORS

Dear MJM,

I joined the editorial team at MJM in March 2003. Over the last year I have greatly enjoyed being an external editor for MJM and welcome the opportunity to get involved with the production of such a high quality journal. My duties as an external editor primarily involve reading and commenting on papers submitted to MJM. However, as the number of papers that I have been involved with accumulates, I have noticed that authors tend to make a number of common errors. In general these are errors in presentation and style, rather than science. I have summarised these here as 12 tips for potential authors.

1. Above all, remember that the editors are busy people. Although we all try to be as objective and constructive as possible, like examiners and members of a job interview panel, we will be influenced by first impressions. Pay attention to details and don't let poor presentation let you down.

2. Don't be afraid to ask for help. Writing a paper for publication is a skill that can take years to develop. Many faculty members will be more than impressed that you are even attempting it and more than willing to offer advice on style, presentation and content.

3. Read the "Instructions for Authors" on MJM's website (http://www.mjm.mcgill.ca). These detail clear directions concerning the presentation and structure of manuscripts, manuscript length and the citation of references. If you do not stick to these instructions it gives the editors the impression that you do not care much about your work.

4. Be clear what your aim is. Decide what you wish your aim to be and state this unambiguously as early in the paper as possible. When you have finished check that you have achieved your stated aim and that this is clear from your conclusions.

5. Use headings and subheadings - even if the instructions for authors do not require these in the type of article you are writing. Headings help to structure your thoughts and, therefore, your writing. They help any argument you are trying to make flow in a logical order and should help you avoid repeating yourself.

6. Keep your writing specific and simple. Go through every sentence and ask: what do I mean by this sentence? Have I said exactly what I mean? Is there any way it could be misinterpreted? Are there any unnecessary words? This process can be time consuming until you are used to it, but the result should be a much better piece of writing which requires much fewer revisions if accepted.

7. Remember that MJM is an international journal. Don't assume that all your readers will be Canadian or even North American. If you are referring to some procedure or policy that is particular to Canada you must explain it for someone who has no knowledge of it.

8. Remember that MJM is a general medical journal. If you are submitting a paper on a particularly specialised topic, you must be careful to explain it for a general audience who have a broad knowledge of medicine but not necessarily specific knowledge of the topic you are discussing.

9. If you are reporting original data, remember that data description is not the same as data analysis. Provide an appropriate statistical analysis of your data and make sure you are clear what the results of your statistical tests mean.

10. Never confuse association with causation. A relationship between two variables does not mean that one causes the other, it merely means that the two are associated with each other. Association is a necessary, but not sufficient, criterion for establishing a causal relationship between two variables.

11. Remember that medicine and biomedical sciences are complex and evolving fields in which there are always controversies and differences of opinion. Learn how to search the literature and critically appraise the papers that you find. Make sure the sources you cite are up to date and that you do not rely on a limited number of sources.

12. Always proofread your work before submission - and then proofread again. Many papers submitted have numerous basic spelling and grammar mistakes which should have been spotted before submission. If you cannot be bothered to read your work why should the editors or readers of the journal? If English is not your first language, ask someone who is fluent in English to help you check your work before submission.

I hope that these points will be of use to authors considering submitting manuscripts to MJM.

Yours faithfully,

Jean Adams, BMedSci, MBBS, PhD (2004)
School of Population and Health Sciences
University of Newcastle upon Tyne, UK

Jean Adams is in the final year of a three year PhD research fellowship. She obatined her medical degree in 2001 and also has a BMedSci degree in health psychology and psychiatry. Her current research focuses on socio-economic inequalities in health.

COMMENTARIES

ELECTIVES IN THE DEVELOPING WORLD

Choosing to do an elective in a developing country can be a gamble. Everyone has heard stories of unsuspecting students being deposited in a remote hospital and told that actually they are the doctor, and here are the hordes of patients they must look after for the next eight weeks. As if this isn't enough, the students often find themselves suffering severe culture shock, working in a health care service that bears no resemblance to the one they are accustomed to and encountering patients who frequently present with either unfamiliar diseases or unfamiliar presentations of familiar diseases.

Some would argue that this is an exciting and worthwhile experience from which the adventurous student can learn vast amounts. However, others may find it a lonely and stressful two months in which there is a lot of potential to blunder into problems through lack of understanding of local culture and health care practices. Worse still, for many students the prospect of working in a developing country is too daunting to even be considered an option.

The purpose of this article is to remove some of the fear from doing an elective in the developing world. Examples of the author's own experiences in Tanzania are used, and a few simple guidelines are provided which students can follow to ensure that they are fully prepared; enabling them to get the most out of their elective, and also give the best back. Finally, the concept of an "International Health Elective" as a model for successful electives in the developing world is discussed.

How to prepare for your elective

Decide where you want to go!

Obvious, but it is important that you make the right decision. A good resource to help you make this decision is other students who have completed their elective. You can usually contact such students through your medical school. Useful information can frequently be found in student medical journals, many of which carry students' elective reports from various countries(1). There are also books available for medics considering working abroad(2), and websites that cover international health issues(3). If concerns about safety may influence your decision, contact the foreign office in your country for some advice(4).

Red Tape and Practicalities

It is important to get the practicalities out of the way once you have chosen a destination and arranged your dates of travel. It is best to do this at least 3-6 months before you travel, as vaccination schedules may require a few months between boosters. Most of this information is available in good guide books (see Resources section at the end), but some key points relevant to medical electives are set out below.

Your health when abroad is extremely important. Make sure that you visit a travel clinic (either within your University or with your own doctor's practice) to obtain advice about necessary immunisations, appropriate anti-malarial prophylaxis and general health and safety issues relevant to your chosen destination. It is also worthwhile obtaining a regime of HIV post-exposure prophylaxis drugs if the country you are visiting has a high prevalence of HIV; your University may be able to arrange this for you. Should you become unwell, it is useful to have a list of generic drug names so that you can easily request the medications you need. Some people even carry a kit with a few basic medications (analgesics, anti-diarrheals, re-hydration salts, antibiotics such as ciprofloxacin and metronidazole, and anti-emetics). However, this is not a substitute for medical advice if you are unfortunate enough to become ill while on your elective. Importantly, you must ensure that you have good health insurance for your elective - it is reassuring to know that your insurers will pay to fly you to hospital should you be struck down by a case of cerebral malaria in the middle of the African countryside. Additionally, you must ensure that you have insurance that covers you as a health care provider in your chosen elective destination. Check with the medical insurance groups in your home country as many of them also cover elective students overseas.

Finally, financing the elective. Depending on your destination, the costs of flights and living expenses may not be excessive, but can nevertheless be significant. It is worthwhile to look for grants and bursaries that will offset at least some of the costs of your trip. Often grants are given if you are intending to carry out research whilst on your elective (for example, those from the Wellcome trust(4)). Most grants are specific to the types of research or country being visited. Your medical school will be able to tell you which grants and bursaries are available to you and how to apply for them.

Learn about your destination country

This greatly enriches the elective experience. Developing countries are by their nature very different from countries in the west; it helps to understand these differences before you go. Find out about the economic

status, political climate and culture of the country you are going to visit. Try and study its history, how it became part of the "developing world" and what this label means for it now. Doing this will help you understand the behaviour, attitudes and lifestyle of the people around you, making you less prone to culture shock. If you have time, read about the global players that influence the politics (and thus health care provision) in the country, such as the World Bank, International Monetary Fund and World Health Organisation. Some useful references are listed in the resources section below.

It also pays to "know the lingo"; essential for communicating with local people within the hospital and the community. Make sure you can speak at least one of the national languages. Whilst you can often get by with English in many developing countries, it helps if you have taken a few classes in local languages before you arrive-particularly when it comes to getting histories from your patients.

Learn about the local healthcare system

The government health care system in many developing countries is based on a western model that has evolved according to the changing political climate. Alongside this many countries have a private health care service and traditional healers. Before you go, find out about the health care system that exists in your destination, what services are available (public/private/traditional), how these are financed, who uses them and how effective they are at meeting the health needs of the population. This will help you understand the circumstances under which the medical staff must work; having some knowledge of how desperately under-financed the Tanzanian health care system has certainly helped us to appreciate the stress and attitudes of the staff we were working alongside.

Find out about the general health of the population in the country you are visiting. Which diseases are common and how do they manifest themselves? What is the life expectancy? What is the HIV prevalence rate? Statistics can be found on some of the websites listed at the end of this article.

Contact the hospital

Make contact with the appropriate people at your chosen destination and attempt to organise a programme of study before you go. Arrange to spend time learning alongside students wherever possible. Ensure that your experience will be in all areas of the health care system (public/private, rural/urban, primary/tertiary) as this will give you a fuller understanding of the health care situation. Find out which medical staff are in the hospital and try to arrange a senior doctor to be your supervisor when you are in the country, with whom you should agree on objectives you want to achieve during your elective. Your supervisor should be able to facilitate if things are not going according to the programme.

In Tanzania it was possible to arrange a comprehensive programme of study before we left the UK. We were allocated to specific medical firms, and a week was set-aside for us to accompany a group of students on their community placement to a rural hospital. This thorough preparation meant that we were able to see how health care was provided at all levels during our two-month stay. The rural placement allowed us to see for ourselves the difficulties faced by the majority of Tanzanians: an invaluable and humbling experience.

Make contacts before you go!

As mentioned, one of the most important resources you can use when preparing for your elective is information from medical students in the country you are planning to visit. You can do this using international student networks such as the International Federation of Medical Students' Associations (IFMSA). Students will let you know what to expect when you arrive in the country, tell you the sorts of things you will need to bring, and of course it is advantageous to build up friendships before you arrive.

When you arrive

Students are an even greater resource once you are on your elective. They know a lot about the conditions you will see, and also the predicament of doctors within that country. They can provide the support, assistance and understanding of issues (both inside and outside the hospital) needed to get the most out of the elective experience. Spending time with them will help you understand important cultural differences between your home country and elective destination, such as time keeping, behaviour and dress code. While you may find some of these differences difficult to accept, it is necessary to respect, and avoid offending local people.

Working with other students on the wards and in clinics helps to overcome language and cultural barriers. In Tanzania, we frequently relied upon students to assist in translating histories. The students also have plenty to gain from you; the exchange of skills, knowledge and attitudes that occurs as you interact is of mutual benefit. We demonstrated how to examine a patient with suspected Parkinson's disease (a rare condition in Tanzania), and they taught us about malaria, HIV, tuberculosis and other diseases that they commonly encountered. Ultimately, it was the close friendships that we built up with the medical

students in Tanzania that truly enhanced our elective experience.

When you return home

Keep in touch with your peers in the country you have visited. Friendships between students facilitate the future exchange of knowledge, ideas and understanding between the two countries. In an increasingly globalised world this sort of communication helps you to keep up to date with what is happening to health care on a global scene.

Final points on completing an international health elective

The International Health and Medical Education Centre (IHMEC) at University College London runs an international health elective programme. Students first complete four weeks of pre-elective tuition covering issues that affect health care provision in the developing world, preparing the ground for the elective. When abroad, they are tutored by local health care workers and faculty teachers at all levels of health care, in conjunction with the local students. Students benefit from being in close contact with their peers in the host country, and also feel reassured by the integrated program and support provided. This year is the third year that IHMEC has run international health electives and students have been placed in several different developing countries.

Not all medical schools have an international health centre within their university, but students across the globe are free to use IHMEC as a resource to assist them in preparing for their elective experience. The centre is currently in the process of designing a pre-elective pack for those who cannot do the full elective programme of tuition prior to travel and de-briefing on returning home. This covers issues of globalisation, poverty, donor aid, access to drugs and other topics that affect health care provision in the developing world. It also includes a list of useful reading material and advice (3).

Conclusion

It is important to be fully prepared for your elective, particularly if you are aiming to visit the developing world where lifestyle, customs and living conditions are likely to differ significantly from those that you are used to. Whether just by taking a little time to do some reading before you go or experiencing an "international health elective", you can enhance your experience, making your elective both educational, enjoyable and of benefit to both you and your hosts.

Resources

General information

1.International Health and Medical Education Centre: www.ihmec.ucl.ac.uk
2.CIA: www.cia.gov/cia/publications/factbook
3.UK Foreign and Commonwealth Office, for general information (and any travel warnings) about specific countries: www.fco.gov.uk

Your health

1. UK Department of Health travel advice: www.doh.gov.uk/traveladvice
2. London School of Hygiene and Tropical Medicine: www.lshtm.ac.uk
3. Centers for Disease Control and Prevention (information about HIV post-exposure prophylaxis: www.cdc.gov/travel
4. Jones N. The Rough Guide to Travel Health. Rough Guides, 2001

Background reading

1. The Rough Guide and Lonely Planet guidebooks, including the Lonely Planet website: www.lonelyplanet.com Wilson, M. The Medics' Guide to Work and Electives Around the World (2nd Edition). Edward Arnold, London, 2003.
2 Jong EC, McMullen RM. Travel and Tropical Medicine. WB Saunders, 2002, Eddleston M, Pierini S. The Oxford Hnadbook of Tropical Medicine. Oxford University Press, 1999.
3. Bell DR, Beeching NJ, Gill GV. Lecture Notes on Tropical Medicine (5th edition). Blackwell Science UK, 2003.

Health care in the developing world

1. World Health Organisation: www.who.int
2. International Committee of the Red Cross: www.icrc. org
3. Médecins sans Frontiéres: www.msf.org
4. Centers for Disease Control and Prevention: www .cdc. gov
5. International AIDS Vaccines Initiative: www.iavi.org
6. Global Fund for AIDS, Tuberculosis and Malaria: www.globalfundatm.org

Developing world economics

1. United Nations Development Programme : www.undp.org
2. World Bank: www.worldbank.org
3. World Trade Organisation: www.wto.org
4. World Development Movement: www.wdm.org.uk
5. Hunger notes (food security issues): www.worldhunger.org
6. MedAct (an NGO for health professionals): www.medact.org

7. One World: www.oneworld.net
8. Unicef: www.unicef.org
9. The United Nations: www.un.org
10. Millenium Development Goals: www.developmentgoals.org

Electives in the developing world

1. The Student BMJ website (www.studentbmj.co.uk) carries an elective report in all of it's issues, frequently from developing countries. Search the website for a report from any country you are particularly interested in.
2. Banatvala N, Doyal L. Knowing when to say "no" on the student elective. Students going on electives abroad need clinical guidelines. British Medical Journal. 1998 May 9;316(7142):1404-5

Getting in touch with students

1. International Federation of Medical Students Associations: www.ifmsa.org

REFERENCES

1. The Student British Medical Journal: www.studentbmj.co.uk
2. Wilson, M. The Medics' Guide to Work and Electives Around the World (2ndEdition). Edward Arnold, London, 2003.
3. International Health and Medical Education Centre: www.ihmec.ucl.ac.uk
4. The Wellcome Trust: www.wellcome.ac.uk

Dr. Shiner is currently working as a pre-registration medical house officer at the Royal Preston Hospital, Lancashire, UK. She is interested in pursuing a career in international public health policy. Dr Jaime Miranda is the International Health Electives Co-ordinator at University College London

ORIGINAL ARTICLE

A Retrospective Study of the Management of HIV, Hepatitis B and Hepatitis C-Positive Pregnancies in Edinburgh, UK from 1997-2002.

Wendy J. Russell *.

STUDY AIMS: This study aims to examine management practices for HIV-positive, HBV-positive and HCV-positive pregnancies over 1997-2002 in Edinburgh, UK, and the effects the diseases have on pregnancy outcomes. RESULTS: Equally for HIV, HBV, and HCV, 50% of the diagnoses were made before pregnancy while the other 50% were detected and diagnosed through antenatal testing. Of the 17 HBV-positive pregnancies 31.6% of the women were highly infectious at delivery and 57.9% were carriers with low infectivity. Of the 17 HIV-positive pregnancies 47.1% of the women had an undetectable viral load and 17.6% were unrecorded at delivery. All 17 HIV-positive pregnancies received ART in varying regimes, 15 (88.2%) were on combination therapy, one delivered vaginally and no women breastfed. All neonates of HBV-positive mothers received immunoglobulin and vaccination and were then breastfed. There were no specific interventions for HCV. Only one study child out of the 38 pregnancies became infected, and this was with HIV. CONCLUSION: Routine screening identifies women with no obvious risk factors, and interventions are largely accepted and effective at reducing vertical transmission. HIV therapy is individually tailored and increasingly uses several agents. Moreover, there is a movement towards allowing low viral load HIV-positive women to deliver vaginally. There are no interventions recommended for HCV infectivity alone. The difficulty collecting information illustrates that no adequate tracking system of infected pregnant women exists. Recommended is the creation of a formal database that includes standardized information such as the viral load of HIV or HCV at delivery, so that outcomes of intrapartum management can be more effectively assessed. No comment can be made on virus-related pregnancy complications, as study numbers are too small for statistically valid data.

KEY WORDS: Human Immunodeficiency Virus, Hepatitis C, Hepatitis B, Pregnancy, Disease Management, Antiretroviral Therapy

INTRODUCTION

Human Immunodeficiency Virus (HIV), hepatitis B (HBV) and hepatitis C (HCV) are important blood-borne viruses with significant morbidity and mortality and proven vertical transmission(1,2,3). Until recently HIV and HBV were only tested for in those patients perceived to be at risk of exposure. However, with growing understanding of how and when vertical transmission occurs, and proven interventions to reduce this, routine optional antenatal screening was introduced in 1999 to identify infected women with no obvious risk factors.

* To whom correspondence should be addressed: 11 Sciennes Gdns, Edinburgh, EH9 1NR. Email: wends@doctors.org.uk

Human Immunodeficiency Virus

HIV has a high morbidity and mortality resulting in significant treatment and care costs. Although incurable, antiretroviral agents can increase the life expectancy of those infected by slowing viral replication and lowering viral load in the body before irreversible damage to the immune system occurs (4). Via this mechanism they then reduce the chances of the transmission to a child of an infected woman during pregnancy or labour. However, high viral turnover and mutation rates lead to drug resistance and therefore the use of several drug categories is becoming common practice. As resistance grows within the HIV infected population, methods to reduce transmission are becoming ever more important.

Globally women account for 47% of those living with HIV,5 and they are becoming increasingly infected. The vertical transmission rate is 15-40% (6),and is influenced by preterm delivery (7), prolonged membrane rupture (>4 hours) (7), advanced disease, high viral loads and breastfeeding (1,8,9).

Antenatal diagnosis allows time for management planning producing a greater chance of undetectable viral loads at delivery, and identifies infected women without obvious risk factors before symptomatic presentation. Each year over 300 HIV infected women give birth in the United Kingdom, two thirds of which are in London(8). This produces an HIV infection rate of <3/10,000 pregnancies(9).

Children who become infected during birth progress to symptomatic HIV and AIDS at a faster rate without treatment (8-10years) (1). However, universal screening allows at-risk children to be followed up for early diagnosis and treatment, providing a greater prospect of years of healthy normal development (2).

In 1994 the first regime to reduce vertical transmission was developed. The aim was to lower maternal viral load, avoid neonatal contact with infected maternal fluids and treat any transmitted virus. Large multi-centre, randomised, placebo-controlled trials of antiretroviral therapy (ART) and several meta-analyses of large prospective studies into mode of delivery show that antenatal, intrapartum and neonatal ART, a Caesarean delivery and abstinence from breastfeeding reduces the vertical transmission rates to <5% (8,9,10,11,12). Due to the success of this management, routine screening was introduced in Scotland with an aim to get 90% of pregnant women to accept the tests to therefore identify 80% of HIV infected women by mid-2003 (6).

When diagnosing HIV in the child of an infected mother, the standard enzyme linked immunosorbent assay (ELISA) test for IgG is not suitable until after 18 months of age due to transplacental passage of maternal antibodies. Therefore a polymerase chain reaction (PCR) for viral RNA is used in the early months alongside health checks in follow-up. A confidential voluntary surveillance system of these children is in place over the UK (the Collaborative HIV Paediatric Study of UK and Ireland or CHIPS) from which transmission rates and disease progression statistics are taken.

Hepatitis B Virus

HBV infection can cause serious disease including liver failure, cirrhosis and hepatocellular carcinoma. The infection can be asymptomatic but around 10% of infected adults, and 90% of infected neonates, become long-term carriers with a 25% lifetime risk of cirrhosis (13). Due to rising drug addiction and immigrant populations, chronic carriers are more frequently seen in the UK (14). Co-infection with HIV makes HBV a more severe disease, resulting in prolonged illness, more complications, and a larger number of chronic carriers.

Without treatment, vertical transmission is 10-20% and linked to infectivity (3) and perhaps foetal exposure in threatened abortion or pre-term labour. Immunization is a successful and safe procedure used for many at risk populations throughout the world, and in some areas, is now part of the routine vaccination protocol. For post-exposure protection (including infants born to infected women), vaccination and IgG is available with 90% efficacy if given within 12 hours (13). After this time, breastfeeding is not contraindicated (15) and the general practitioner completes the vaccination programme.

Selective screening of those "at risk" during pregnancy fails to identify 30-50% of infected women, and because the interventions are simple to implement, universal screening has been introduced.

Hepatitis C Virus

Like HBV, HCV can be a serious disease, with up to 80% of those infected becoming chronic carriers and 20% of those developing cirrhosis over 20 years.16 However, unlike HBV, there is no vaccination for HCV. Due to a long asymptomatic incubation, estimating prevalence of HCV in the population is difficult. Available data does suggest that approximately 3% of the world's population is infected with HCV (with about 2.5% of Europeans)(16) and one large Italian study found that 2.4% of pregnant women had anti-HCV antibodies with 72% of them positive for HCV-RNA (17) The group most at risk of contracting HCV are the intravenous drug abusers with up to 90% infected in some studies (16,18), and again, co-infection with HIV worsens the prognosis with more rapid progression to liver failure (19).

Originally HCV was managed with Caesarean section and avoidance of breastfeeding, but recent clinical trials show no benefit from this,18,20,21 and currently the risk of vertical transmission is <5%20. However this is increased if the mother is viraemic for HCV (20,22,23). Therefore, while HCV is readily detectable and arguably prevalent in some areas, the low transmission rates and the lack of effective interventions to reduce the rate further, make routine screening inappropriate at present. Despite this, when infected with HCV and HIV, not only does the risk of HCV transmission double (20), but transmission of HIV increases also (20). This makes HIV management issues even more critical and so, patients "at risk" of having HCV (for example IV drug abusers), are often tested for this virus when pregnant.

AIMS

This article aims to look retrospectively at a cohort of HIV, HBV and HCV-positive pregnancies from 1997 to 2002 in Edinburgh, to study the interventions implemented in their management. Current information about the diseases and recommendations or ideas pertaining to their management will be explored, alongside data collected from the study case records including: timing of diagnosis, uptake of planned interventions, and pregnancy outcomes (post-partum haemorrhage, gestation, birth-weight and infectious status of the child).

METHODS

Inclusion criteria were HIV, HBV and HCV infected women, previously and newly diagnosed, who gave birth to a live infant from 1st September 1997 to 31st August 2002 (a total of 5 years) at the Royal Infirmary of Edinburgh and later, when the hospital moved sites, the New Royal Infirmary of Edinburgh.

Several selection methods were used to find patients. Previously diagnosed women who had antenatal viral loads measured, and patients newly diagnosed in pregnancy were found by a virologist in laboratory antenatal records. The midwife specifically looking after the HIV and HBV-positive pregnancies had created a list of names that was utilised. A paediatrician following up the children of these pregnancies helped the midwife to work back from the children's names she had, to their mother's. Finally, from the women found by the above methods, the computer system in the obstetric department was used to identify any other pregnancies over the selected time period.

The obstetric records of these patients were collected and pre-selected data were gathered from them. Data not included were pregnancies resulting in miscarriage or termination, women who moved out of Edinburgh to deliver, notes that could not be found and women who could not be traced from their child's name (often the family name would be changed after delivery). Many of the HCV cases relied on this trace-back method, as there is no other database of affected pregnancies, and thus, many HCV affected pregnancies could not be identified for inclusion. However, most of the known HIV and HBV-positive pregnancies were included in the study.

RESULTS

Study Groups

Table 1 displays the numbers of pregnancies, number of affected women and numbers of deliveries for each viral group.

Timing of diagnosis

Each pregnancy of each woman was counted as an

Table 1. Number of pregnancies in each disease category

Disease	Number of pegnancies	Number of women	Number of sets of twins	Infants delivered
HIV-positive (total)	17	16	2	19
HBV-positive (total)	17	15	0	17
HCV-positive (total)	7	7	0	7
HIV-, HBV-positive	0	0	0	0
HIV-, HCV-positive	3	3	0	3
HBV-, HCV-positive	0	0	0	0
HIV-, HBV- & HCV-positive	0	0	0	0
Total	38	35	2	40

1. Twin pregnancies were counted as one pregnancy but two infants.
2. Three women had more than one pregnancy during the time period: each pregnancy counted as a separate event but the women were only counted once.
3. As women coinfected with HIV and HCV were counted in both the HIV (total) and HCV (total) groups;
the overall total number of pregnancies = HIV(total)+HBV(total)+HCV(total)-HIVandHCV.
4. Two HIV-positive women were Hep B core antibody positive but Hep B surface antigen negative (therefore having resolved HBV) and so were classed as HIV-positive only.
5. One woman with HIV also had HCV but, even though she was not viraemic at delivery, she was included in the co-infected HIV and HCV group.

individual event. For each virus, approximately half of the cases, 9 of 17 HIV cases, 7 of 17 HBV cases and 4 of 7 HCV cases, were prenatal diagnoses and the remaining, with the exception of one case of HBV, were picked up during antenatal testing. The diagnosis of this case of HBV was only made after delivery when a paediatrician questioned the need for neonatal vaccination for HBV due to the mother's country of origin (Senegal).

Interventions for HIV infectivity.

No two pregnancies to HIV-positive women were managed exactly the same way (Table 2). All women received antenatal antiretroviral therapy, however, the combinations of agents used and the timing of its initiation varied. Also only 35.5% received the recommended zidovudine infusion during labour. Reasons for these treatment variations included previous drug regimens of the women involved, their resistance patterns and their viral loads during pregnancy.

For 15 of the 19 infants born to infected women, treatment was recorded, and the most commonly used agent was zidovudine (Table 2). However, prescription length differed from three to six weeks, and combination therapy was used occasionally.

Overall, ART recommendations for HIV-positive pregnancies were followed but with individual tailoring, resulting in 15 (88.2%) of the 17 pregnancies receiving combination therapy.

All but one of the HIV women delivered by Caesarean section (Table 3): 12 were elective because of the infection, with or without a history of previous sections, and 4 required an emergency procedure for spontaneous labour, twins, or placenta praevia.

Prolonged membrane rupture has been linked to increased rates of vertical transmission, however, with limited patient numbers, there is difficulty calculating averages for this. Following recommendations, none of the HIV-positive women breast-fed.

Table 3. Delivery Routes

Delivery Route	HIV-positive (total)	HBV-positive (total)	HCV-positive (total)	HIV-, HCV positive (total)
Spontaneous Vertex	1 (5.9%)	9 (52.9%)	3 (42.9%)	0
Forceps	0	3 (17.6%)	0	0
Elective Caesarean	12 (70.6%)	2(11.8%)	2 (28.6%)	2 (66.7%)
Emergency Caesarean	4 (23.5%)	3 (17.6%)	2 (28.6%)	1 (33.3%)
Total Number of deliveries	17	17	7	3

1. Pregnancies to women co-infected with HIV and HCV were counted in [HIV-positive (total)], [HCV-positive (total)] and [HIV and HCV-positive (total)] categories

2. As there were no women with HIV plus HBV or HBV plus HCV or HIV plus HBV plus HCV, these categories are omitted from the table

Interventions for Hepatitis B infectivity.

The majority (70.5%) of deliveries for HBV-positive women were vaginal (Table 3). Reasons recorded to explain the use of Caesarean sections were mainly obstetric, however, in contrast to recommendations at the time, one woman, with low infectivity, was stated to have had a section partly because of her HBV status.

Following recommended protocol, all neonates received immunoglobulin and the initial vaccination dose at birth, and all then went on to be breastfed. The remaining vaccination doses were to be given at one and six months by the child's general practitioner.

Interventions for Hepatitis C infectivity

Of the 4 women infected solely with HCV, 3

Table 2: Antiretroviral therapy for HIV

	During Pregnancy			During Labour	Neonatal	
	MonoTherapy	Combination Therapy	IV infusion	Mono-therapy	Combination Therapy	Undiscolsed
Number (% of total)	2(11.8)	15 (88.2)	6 (35.5)	9 (47.4)	6 (31.6)	4 (21.1%)
Drugs used (descending order of use)	Zidovudine	Zidovudine Lamivudine Nevirapine Nelfinavir Abacavir Stavudine Didanosine	Zidovudine	Zidovudine Nevirapine	Zidovudine Nevirapine Lamivudine Abacavir Nelfinavir	

delivered vaginally (Table 3). The reason stated for the use of Caesarean for the other delivery was based on obstetric as opposed to infectious grounds. The women co-infected with HIV and HCV all received Caesareans. Women with HCV only were advised to breastfeed, and for reasons not stated, only one of the 4 decided to do so.

Infectivity of the mother at delivery

Vertical transmission risks are associated with a high HIV or HCV viral load or "highly infectious" HBV (3,9,18). Highly infectious HBV is diagnosed by a positive titre for hepatitis B-e antigen (HBeAg) and a negative titre for the antibody to the e antigen (HBeAb), therefore the virus is actively replicating without an adequate immune response. Measuring viral load at delivery can also be used to assess the efficacy of the antenatal management of HIV.

The aim of antiretroviral therapy is to achieve an undetectable viral load at delivery (<50 copies/ml) and this was achieved in 8 of the 17 pregnancies (47.1%). Unfortunately, three of the deliveries had no viral load recorded, and ART did not achieve maximal effect for one woman who delivered with a high viral titre of 254,000 copies/ml. The numbers are too small to know if the timing of treatment initiation influences viral load at delivery.

For HBV 6 of the 17 deliveries (31.6%) were highly infectious at delivery (see above for criteria), and the remaining 11 (57.9%) were carriers with low infectivity (HBeAg negative and HbeAb positive). The criteria for having resolved HBV is a positive titre for hepatitis B core antibody (HbcAb) and a negative titre for hepatitis B surface antigen (HBsAg) showing previous infection. Two women had resolved HBV but because they were not infectious, they were not included in the HBV infected study group.

There is no routine measurement of HCV viral load, and therefore the data could not be included.

Infective status of the child

Of the 19 children born to HIV-positive women in this study, 7 currently have indeterminate infectious status (too young at the study conclusion to reliably determine a diagnosis), 11 are HIV-negative (antibody negative on at least two occasions) and only one was found to be infected in follow-up but is asymptomatic on ART at 18 months (Table 4). It was not possible at the time of writing this article to determine the delivery route of this child.

As the General Practitioner follows-up children born to HBV-positive women, the precise number of transmissions is unknown; however, from the sources approached there are no known cases.

Children born to mothers infected with HCV are followed up by the same paediatrician as the HIV cases, but receive no interventions. There are no reported neonates with HCV.

Table 4. Infectious status of the children

Status	HIV	HBV	HCV
Positive	1 (5.4%)	0	0
Negative	11 (57.9%)	17 (100%)	17 (100%)
Indeterminate	7 (36.8%)	0	0
Total	19	17	7

Table 5. Pregnancy Outcomes

Outcome	HIV-positive only	HBV-positive only	HCV-positive only	HIV- and HCV positive
Average delivery gestation (wks)	36	39	36	38
Numbers preterm (<37 weeks)	5 NB includes 4 twins	3	1	0
Average blood loss for delivery (ml)	439	415	288	483
Number with PPH (postpartum heamorrhage) (>500ml)	2	3	0	1
Number with significant PPH (>1000ml)	1	1	0	1
Average birthweight (g)	2644 (3589 without twins)	3266	2258	2697
Number with low birth weight (<2500g at term)	4 NB all twins	0	2	1
Number of infants requiring time in neonatal unit and why	7 =4 twins +1 premature +2 withdrawal	3 =2 preterm +1 meconium	1 =1 preterm	1 =1

1. For HIV there were 17 pregnancies, resulting in 19 infants (ie 2 twin pregnancies).
2. Multiple pregnancies result in earlier delivery and smaller infants.

Effect on pregnancy

On planning this study it was hoped that information about the effect of these viruses on various pregnancy outcomes could be assessed, however it soon became apparent that study numbers were not large enough to take into account confounding factors and calculate valid statistical data. The averages for the measured outcomes are included in Table 5 for interest.

Overall, average gestation and birth weight for non-twin pregnancies were adequate and no major adverse outcomes linked to the viruses could be noted.

DISCUSSION

The aims of this article were to look back over pregnancies affected by HIV, HBV and HCV (or any combination) from September 1997 to August 2002 and study their management. It also hoped to look at the effects these viruses had on various pregnancy outcomes and review current understanding about recommendations for the management of these virus.

Human Immunodeficiency Virus

Because HIV is currently incurable, reducing the spread is a vital part of management on a global scale. Vertical transmission from mother to unborn child is one area that can readily be addressed with effective interventions(8,9,10,11,12). Because of this, universal antenatal screening for HIV was recently introduced with an aim to identify previously unrecognised infected pregnancies and allow appropriate management(6). From the start of September 1997 until the end of August 2002 informed estimates put the number of deliveries in Edinburgh at around 32,000. From these, 17 HIV infected pregnancies to 16 women that resulted in 19 live infants were identified. Around half of these pregnancies were known about before the booking appointment for antenatal care, and the remainder were detected via the routine antenatal screening. It is encouraging to see that the HIV screening is working in its aim to identify previously unknown infections.

While there was no clear pattern to the choice of antiretroviral agent or when it was initiated, there was 100% uptake of ART in some form for the pregnancies. There are many reasons as to why therapy differed from patient to patient. Historically zidovudine (a nucleoside reverse transcriptase inhibitor or NRTI) was the first anti-HIV drug to be introduced. Initially it was used as monotherapy in the management of HIV in pregnancy and is still used as an infusion during labour to reduce intrapartum transmission(7), and for neonatal treatment after delivery. But due to the high replication and mutation rates of HIV, resistance is a growing concern. Because of this, combination ART is increasingly being used. By utilising agents with different sites of action (often by using two NRTIs with or without a Protease inhibitor or non-NRTI1) there is less selection pressure for a resistant strain and control is more effective. The choice of which agents to use within a combination regimen depend on the resistance patterns of the virus within each patient, the tolerability of the side effects produced and the pattern of interactions between the drugs. For these reasons, the ART regimen must be individualised for each patient, and revised if necessary throughout treatment.

In this study 88.2% of the pregnant women received combination ART which is in line with current recommendations(1), and because of this, nearly half of the deliveries were with an undetectable viral load.

As the use of any drug during pregnancy raises concern about the safety to the developing unborn child, there have been studies into the outcomes of in utero exposure to ART. Although more research into the long-term effects of ART is needed, two large multi-centre, randomised, double-blind, placebo-controlled cohorts of in utero exposure to zidovudine and combination therapy, found no associated risk of adverse short or long term outcomes(24,25).

Due to the efficacy of combination therapy, which results in so many women with the ideal undetectable viral load at delivery, there is now increasing discussion about the need for Caesarean sections to avoid infant contact with maternal fluids. By avoiding surgical intervention, risks and recovery times for the mother are reduced and she also gains more choice and empowerment over the birth of her child. While most women in this study received either an elective or emergency section, there was one mother who was allowed a vaginal delivery because she had undetectable virus. This may reflect the transfer of these academic discussions into clinical practice, however there are no clear studies to support the safety of these proposals yet.

While the avoidance of breastfeeding may not be appropriate in developing countries where the availability of safe water and affordable formula feed mean that mortality and morbidity from infection outweighs the benefits of avoiding HIV contraction, in the UK this is generally not the case. All HIV-positive women should be advised not to breastfeed their child, despite receiving neonatal ART. In this study, no women breastfed.

Very encouragingly there was only one reported case of vertical transmission over the 5 years, and because the infective status of the mother was known, the child was diagnosed asymptomatically thus allowing optimal ART management, resulting in a healthy, active child at 18 months of age.

In summary, the current antenatal screening is effective in detecting unrecognised infected women. While treatment was individualised, the management recommendations of antenatal and neonatal ART, a caesarean delivery and the avoidance of breastfeeding were followed in this study. Because of these measures there was only one reported case of vertical transmission, and the child remained well on ART.

Hepatits B Virus

HBV can be a serious condition with risks of cirrhosis and hepatic failure. There is a 10-20% risk of vertical transmission(3) which can be virtually eliminated by passive-active immunization (administering a dose of IgG and following the vaccination schedule) within 12 hours of delivery(13). Because of this, HBV testing is part of the antenatal screening protocol.

There were 17 pregnancies to 15 women producing 17 live infants in Edinburgh over the study period. Over half of the pregnancies were diagnosed by screening, and there was 100% uptake of the interventions offered up to the initial vaccination. This shows that, like HIV screening, testing for antenatal HBV is effective in detecting unidentified infected pregnancies. The safe and effective interventions are readily accepted and although there is no central follow-up of these children, there have been no reported cases of vertical transmission in Edinburgh over the study's duration.

Hepatitis C Virus

HCV can also be a serious condition but unlike HBV there is no vaccination currently available. The major risk factor for its contraction is intravenous drug abuse, and its prevalence in that population may be high(16,18). Due to a long asymptomatic period, calculating the overall prevalence is difficult, but estimates put 2.5% of the European population infected(16).

Risk factors for paediatric infection include mother-to-child transmission and transfusion or transplantation of infected fluids or tissues(2,16). Since the routine screening of blood products and organs for HCV in 1991(2), the mother-to-child route will become an ever more significant risk factor. Currently the vertical transmission risk is <5%(20), but that is increased in proportion to levels of viraemia at delivery(20,22, 23).However, as caesarean delivery and avoiding breastfeeding does not reduce this risk(18,20,21) routine screening is not recommended.

One amendment to this recommendation is co-infection with HIV. Not only does the rate of HCV transmission increase, but also the risk of contracting HIV is raised(20). This means that when HIV-positive women are at risk of having HCV, testing is recommended as the HIV management issues become even more important.

This study found only 7 pregnancies to 7 women producing 7 live singleton infants, which is probably a small proportion of the actual number of HCV infected pregnancies over that time period as cases could only be selected from women already diagnosed or those deemed as at risk, who therefore received a HCV test during their antenatal care. The other route used to find candidates was via the children born to infected mothers who were known to a specialist paediatrician. This route was not productive however as the list did not include the mother's name, and often the children would have changed family names since delivery.

As a caesarean delivery is not recommended for HCV infection alone, it is encouraging to see that no woman was given a surgical delivery for HCV infectious status alone.

One aim of the study was to assess the infectivity of the mother at delivery, but as there was no standard viral load documentation this was not possible. Although useful for research purposes, routine HCV viral loads for HCV-only infected women would not change the management of the pregnancy. Therefore HCV viral load testing cannot be recommended as part of the antenatal care package.

There were no reported cases of vertical transmission in the children born to the women of this study.

Secondary Findings

Due to small study numbers, it was not possible to account for confounding influences on pregnancy outcomes and therefore it was inappropriate to take statistical data from these results. In the developing world, intrauterine HIV is associated with low birth weight and prematurity (14), however, averages for gestation and birth weight were not below the normal ranges in this study. In a developed country such as the United Kingdom this may be due to several factors: the high level of medical input resulting in low viral loads; the management of infections related to HIV; good nutrition during pregnancy or a combination of these. Despite not being able to acquire statistical data for pregnancy outcomes, a number of interesting secondary findings became apparent.

Firstly, the pool of knowledge about these viruses is continually growing. As such, the understanding of when and how transmission occurs, and what interventions influence this, result in recurrent updating of management recommendations.

Secondly, the difficulties met trying to find cases to include in this study illustrate that there is currently no adequate tracking system of HIV, HBV and HCV infected pregnancies in Edinburgh. This will be due, in

part, to the low prevalence of these conditions in the population, unlike areas such as London where HIV, especially, is much more frequent.

Despite having several limitations, this study does illustrate important points in management and areas for further research such as vaginal delivery in HIV and long-term antenatal ART. Where these infections are less common, awareness should be raised about the need for service organisation to allow evaluation and improvement.

ACKNOWLEDGEMENTS

Many thanks to Dr K Dundas (Consultant Obstetrician at the New Royal Infirmary of Edinburgh), Dr M Ogilvie (Consultant Virologist at the New Royal Infirmary of Edinburgh), Dr J Mok (Consultant Paediatrician at the Royal Hospital for Sick Children in Edinburgh), Dr Liston (Consultant Obstetrician at the New Royal Infirmary of Edinburgh), Medical Microbiology and Medical Records at the New Royal Infirmary of Edinburgh and Fiona McNeillage (Midwife at the New Royal Infirmary of Edinburgh) for their help finding cases and information and for their support and guidance throughout this project.

REFERENCES

1. Duong T, Ades AE, Gibb DM, Tookey PA, Masters J. Vertical transmission rates for HIV in the British Isles: estimates based on surveillance data. BMJ 1999; 319:1227-29.
2. Gibb DM, Duong T, Tookey PA et al. Decline in mortality, AIDS, and hospital admissions in perinatally HIV-1 infected children in the United Kingdom and Ireland. BMJ 2003; 327(7422):1019.
3. Stevens CE, Neurath RA, Beasley RP, Szmuness W, HBeAg and anti-Hbe detection by radioimmunoassay: correlation with vertical transmission of hepatitis B virus in Taiwan. J Med Virol 1979; 3: 237-41.
4. British National Formulary. Issue 46. September 2003
5. National Institutes of Health Fiscal Year 2003: Plan for HIV related research URL: http://www.nih.gov/od/oar/public/pubs/ fy2003/ i_overview.pdf
6. Scottish Centre for Infection and Environmental Health Weekly Report: 25 June 2002; Volume 36, No. 2002/25. ISSN 1357-4493. p.156.
7. Kuhn L, Steketee RW, Weedon J et al. Distinct risk factors for intrauterine and intrapartum human immunodeficiency virus transmission and consequences for disease progression in infected children. Perinatal AIDS Collaborative Transmission Study. J Infectious Diseases 1999; 179(1): 52-58.
8. Public Health Laboratory Service: HIV and AIDS: Surveillance URL:http://www.phls.org.uk/topics_az/hiv_and_sti/hiv /epidemiology/ quarterly.htm
9. Communicable disease surveillance centre: HIV and AIDS in the UK. An epidemiological review: 2000.
10. Dunn et al. Mode of delivery and vertical transmission of HIV-1: a review of prospective studies. Perinatal AIDS Collaborative Transmission Studies. J Acquir Immune Defic Syndr 1994; 7(10):1064-66.
11. Connor EM et al. Reduction of maternal-infant transmission of Human Immunodeficiency Virus Type 1 with zidovudine treatment. N Engl J Med 1994; 331:1173-80.
12. The International Perinatal HIV Group. The mode of delivery and the risk of vertical transmission of Human Immunodeficiency Virus type 1- A meta-analysis of 15 prospective cohort studies. N Engl J Med 1999; 340: 977-87.
13. Moyes CD, Smith L, Lennon D. Hepatitis B - prevention of perinatal transmission. NZ Med J 2002; 115:19-20.
14. Obstetrics. In: Collier J, Longmore M, Brown TD. Oxford Handbook of Clinical Specialities. 5th ed. Oxford University Press; 2001. p.76-169.
15. Hill JB, Sheffield JS, Kim MJ, Alexander JM, Sercely B, Wendel GD. Risk of hepatitis B transmission in breast-fed infants of chronic hepatitis B carriers. Obstet Gynecol 2002; 99(6): 1049-52.
16. WHO Consultation and the Viral Hepatitis Prevention Board. Global surveillance and control of hepatitis C. J Viral Hepatitis 1999; 6(1): 35-47.
17. Conte D, Fraquelli M, Prati D, Colucci A, Minola E. Prevalence and clinical course of chronic hepatitis C virus (HCV) infection and rate of HCV vertical transmission in a cohort of 15,250 pregnant women. Hepatology 2000; 31(3): 751-5.
18. Garland SM et al. Hepatits C - role of perinatal transmission. Aust NZ Obstet Gynaecol 1998; 38(4): 424-7.
19. Resti M, Azzari C, Bortolotti F. Hepatitis C virus infection in children coinfected with HIV: epidemiology and management. Paediatric Drugs 2002; 4(9): 571-80.
20. European Paediatric Hepatitis C Virus Network. Effects of mode of delivery and infant feeding on the risk of mother-to-child transmission of hepatitis C virus. Brit J Obstet & Gynaecol 2001; 108: 371-7.
21. Tovo PA et al. Increased risk of maternal-infant hepatitis C virus transmission for women coinfected with human immunodeficiency virus type 1. Italian Study Group for HCV Infection in Children. Clin Infect Dis 1997; 25(5): 1121-4.
22. Thomas SL, Newell ML, Peckham CS, Ades AE, Hall AJ. A review of hepatitis C virus (HCV) vertical transmission: risks of transmission to infants born to mothers with and without HCV viraemia or human immunodeficiency virus infection. Int J Epidemiol 1998; 27(1):108.
23. Ohto H et al. Transmission of hepatitis C virus from mothers to infants. N Engl J Med 1994; 330:744-50.
24. Tuomala RE et al. Antiretroviral therapy during pregnancy and the risk of an adverse outcome. N Engl J Med 2002; 346: 1863-70.
25. Culnane M et al. Lack of long-term effects on in utero exposure to zidovudine among uninfected children born to HIV-infected women. JAMA 1999; 281:151-7.

Wendy Russell is a a final year medical student at Edinburgh University, UK.

ORIGINAL ARTICLE

Vagal Tone Biofeedback: Respiratory and Non-respiratory Mediated Modulations of Vagal tone Challenged by Cold Pressor Test

Dorey Glenn, BSc*, Blaine Ditto, PhD

ABSTRACT Deficient parasympathetic activity to the heart has been hypothesized to underlie some forms of cardiovascular disease. Consequently, this experiment attempted to induce non-respiratory mediated increases in cardiac parasympathetic activity. Thirty-nine subjects were asked to increase their vagal tone using biofeedback, paced breathing, biofeedback plus paced breathing, or quiet sitting. The cold pressor test was used to examine the relative efficacy of vagal tone increase to mitigate cardiovascular reactivity. Repeated measures ANOVAs and t-tests revealed significant increases in vagal tone for the paced breathing, and biofeedback plus paced breathing groups, relative to controls. However, the quality of these increases could not be tested using the cold pressor test because the test failed to produce homogeneous cardiac reactivity.

INTRODUCTION

Heart disease and stress-related disorders are among the leading causes of death in the United States and Canada (American Heart Association, 2003; Statistics Canada, 2000). Deficient parasympathetic activity to the heart has been hypothesized to underlie some forms of cardiovascular disease, including cardiac arrhythmias (Eckberg et al., 1971 and 1980). Cardiac parasympathetic activity, or the amount of bioelectrical flow through the vagus nerve (i.e. vagal tone), provides homeostatic and inhibitory influences on the heart (Porges, 1993). Recently, indices of parasympathetic tone to the heart have been shown to reflect overall cardiopulmonary health (23), and thus, methods to increase vagal activity have become of clinical interest. Various psychophysiological, behavioral, and pharmacological methods-including biofeedback-have been devised to increase cardiac vagal activity (22, 13, 26). Vagal tone is most commonly measured by analyzing the rhythmical variations in heart rate period due to the respiratory cycle, generally referred to as the respiratory sinus arrhythmia (RSA) (17).

Previous studies have been able to significantly increase vagal activity (26, 28), but have failed to test these increases with a stressor. Due to the absence of such testing, it is not known whether previous experimental increases in vagal activity are able to mitigate physiological arousal. Acute stressors, such as the cold pressor test, can be used to test the efficacy of

* To whom correspondence should be addressed: Dorey Glenn, 3941 Rivard St., Montreal, Quebec Canada, H2L 4H8. Phone number: (514) 843 7625. E-mail: Dag1111@yahoo.com

training methods and the quality of the resultant increases in vagal tone (30). In addition, past attempts to experimentally increase vagal tone have failed to control for respiratory influences on vagal activity (26). Changes in the respiratory pattern have a profound influence on cardiac variability, and thus readily influence measurements of cardiac parasympathetic activity. The following investigation attempted to use a biofeedback paradigm to isolate non-respiratory increases in vagal tone, and tested the resiliency of these increases with the cold pressor test. Successful increases in vagal tone will not only add support to the role of biofeedback and respiratory control in the treatment of some forms of heart disease, but will shed light on a non-respiratory alternative to vagal tone increases.

Background

The American Heart Association (AHA) estimates that cardiovascular disease costs the United States 351.2 billion dollars in indirect and direct economic losses per year (American Heart Association, 2003). Furthermore, the AHA estimates that cardiac arrhythmias cost Medicare approximately 2.1 billion dollars in 2000. Arrhythmias were mentioned as both the cause of death of 37,646 people, as well as an underlying or contributing cause of death in 461,000 persons in 2002. Due to an increasing proportion of elderly people in both the American and Canadian populations, rates of myocardial infarct and cardiovascular related hospitalizations are expected to rise continuously in the next 20 years (Statistics Canada, 2000). Although cardiovascular disease is a largely preventable pathology, its frequent occurrence has spurned much pathophysiological research, and thus has helped to reveal much of the physiology underlying both normal and abnormal cardiovascular function (18).

The autonomic nervous system has been functionally and anatomically divided into the parasympathetic and sympathetic divisions. The former division is responsible for homeostatic and regenerative functions, and the latter division supports behavioral responses to threat and the mobilization of metabolic resources. The vagus nerve, part of the parasympathetic division, is responsible for much of the neural control of the internal viscera; however, its influence on the heart and lungs is of most concern in the present discussion. Increased parasympathetic activity to the heart decreases heart rate, and in conjunction with respiratory rate and tidal volume, determines respiratory sinus arrhythmia (RSA).

Respiratory sinus arrhythmia is the close occurrence of cyclical fluctuations in heart rate to the respiratory tides. As early as 1733, Hales observed RSA by noticing that changes in pulse were related in a regular manner to the respiratory pattern in the horse. In 1910, Hering had identified the vagus nerve as being responsible for the lowering of heart rate during respiration (23). Since then, computation of respiratory sinus arrhythmias has increasingly been used as a reliable, selective, and non-invasive measure of vagal activity to the heart (5).

RSA is calculated using the slight variation between heartbeat intervals. The presumed purpose of RSA is to maximize oxygen transport during inhalation, and carbon dioxide diffusion during exhalation (22). This physiological mechanism is controlled by the extent of neural activity through the vagus nerve, and can thus serve as a non-invasive measure of parasympathetic tone to the heart (22). Although many factors may produce rhythmical fluctuation in heart rate, the high-frequency component of heart rate variation is principally mediated by the vagus nerve (12). For example: pharmacological parasympathetic blockade (by atropine), but not sympathetic blockade, abolishes respiratory sinus arrhythmia, implicating the vagus nerve as the dominant source of RSA (22,13). Another important factor influencing vagal activity is respiration, which will be discussed in context below.

Parasympathetic tone serves a protective function for the heart, whereby increased parasympathetic tone produces decreased heart rate and increased ability to mitigate sympathetic arousal (23). Porges explains: "Heart rate variability is a marker of the efficiency of neural feedback mechanisms and may index health status or the individual's capacity to organize physiological resources to respond appropriately. Thus, the better the 'organized' physiologic variability, the greater the range of behavior…[and] flexibility in response to environmental demands [or disease states]." In other words, heart rate variability characterizes efficient and sensitive neural control, and thus is a predictor of healthy resilience to stress. Therefore, states of attenuated vagal influence on the heart, manifesting as decreased heart rate variability, are "paralleled by reduced behavioral flexibility in response to environmental demands," and increased susceptibility to a variety a cardiovascular pathologies (23).

Both human and animal studies have revealed a clear association between reduced parasympathetic activity and cardiovascular dysfunction (9, 10). An analysis of the heart rate variability of 808 acute myocardial infarction survivors revealed that heart rate variability was a significant predictor of long-term survival (18). The authors suggest that decreased vagal tone, responsible for decreased heart rate variability, may predispose the heart to ventricular fibrillation.

Likewise, in animals with experimental myocardial infarcts, decreased vagal tone and reduced heart rate variability predisposed the animals to ventricular fibrillation (20).

Furthermore, evidence revealing high correlations between decreased RSA and arteriosclerotic severity (14), and RSA amplitude and the sensitivity of the baroreceptor reflex provide further support of the importance of parasympathetic activity in cardiac health in normal populations (26). Negative correlations between RSA amplitude and blood pressure were found in hypertensive populations (11).

Despite the importance of vagal activity, relatively few non-pharmacological methods to increase vagal tone have been applied at the clinical level (19,28). Among a variety of possible approaches, including autogenic training, progressive relaxation, and transcendental meditation, biofeedback has made important contributions to the field of behavioral modification and psychophysiology (21). Biofeedback is the closure of a loop connecting a physiological variable to a willful individual able to, at some level, control that variable. Although both an awareness model and an operant conditioning model have been proposed as possible mechanisms underlying biofeedback control (6), etiology will not be of concern in this experiment. Biofeedback attempts to mimic the numerous regulatory feedback loops underlying the homeostatic balance essential for life (3). Historically, biofeedback has been used with varying success for a number of clinical pathologies including: hypertension, cardiac arrhythmias, migraine headaches, Raynaud's syndrome, tension headaches, test anxiety, seizure activity, stomach acidity, and incontinence (21).

Many studies have been conducted exploring the efficacy of biofeedback paradigms in manipulating heart rate (25), and to a lesser extent, heart rate variability. One study by Reyes et al. (26) attempted to differentially increase RSA magnitude using variations in biofeedback procedures. Forty subjects were randomly assigned to four conditions: RSA biofeedback without respiratory instructions, RSA biofeedback with respiratory instructions, respiratory biofeedback or, d) respiratory instruction only. Respiratory instructions consisted of slow and deep breathing. Although all groups produced significant increases in RSA from baseline, the respiratory biofeedback group produced the fastest increases in RSA amplitude, and the RSA biofeedback group without respiratory instructions produced the slowest. Between the remaining conditions, respiratory instructions produced faster increases in RSA than did RSA biofeedback with respiratory instructions.

The primacy of respiratory instructions to increase RSA magnitude is not surprising considering that low frequency and high amplitude respiratory patterns induce increases in RSA. However, why should the addition of biofeedback retard RSA increases despite both groups having been given respiratory instructions, as indicated by the latter finding? The authors attribute this peculiar result to "the nature of the target used in each trial." The respiratory instructions group directly manipulated the variable used to control RSA-respiration. The RSA biofeedback group with respiratory instructions attempted to control the variable itself-RSA. The authors view this evidence as reason to focus on: "the control strategies [respiration] [rather] than directly on the physiological variable to be controlled [vagal tone]."

This conclusion may be premature. Increases in RSA were not subjected to a real world "test" or validation; namely, an acute stressor. Thus, one cannot be sure of the quality of these increases. Perhaps respiratory procedures only transiently induce increased RSA, albeit faster. Can these increases withstand or mitigate reactivity to an acute stressor? Does the operant conditioning of biofeedback create a more sustainable form of RSA increase than respiratory instructions alone? Respiratory instructions are cognitive tools subject to degradation and non-use in stressful situations. However, biofeedback learning can produce a pattern or conditioned response to a stressor. In other words, increases in vagal tone must be of good quality, and not necessarily of good quantity. Sakakibara et al. (28) emphasize that "respiratory parameters need to be controlled when assessing the cardiac parasympathetic tone by the magnitude of the respiratory sinus arrhythmia."

The following experiment attempted to separate vagal tone increases due to respiratory manipulations from those increases due to direct, biofeedback-induced, increases in vagal tone. Evidence supporting this approach was found by Sakakibara et al. (1994). During autogenic relaxation, an increase in vagal tone was found even when subjects' respiratory rate and volume were controlled by means of a pneumogram and strain gauge. This showed that vagal tone could be manipulated even when respiration is controlled. Both the mechanism for producing such increases and whether such an effect can be reproduced remains unclear. Nonetheless, Sakakibara et al. (28) hypothesize that enhanced cardiac parasympathetic tone may be one mediator of the relaxation response, as coined and popularized by Benson (4).

Vagal increases due to respiratory instruction or respiratory biofeedback, albeit more rapid, may be as topical and transient as the respiratory cycle itself, and thus susceptible to the vagaries of modern life. It is yet unclear if biofeedback procedures can tap into a

respiratory-contamination-free aspect of vagal tone. By pacing the respiratory rate at a constant rhythm, while administering vagal tone biofeedback, this experiment attempted to produce increases in cardiac parasympathetic tone that were free of respiratory influence.

The cold pressor test was used in conjunction with biofeedback to act as a source of both sympathetic arousal and parasympathetic withdrawal; in other words, as a model for an acute stressor. Past biofeedback research has produced significant increases and mild decreases in heart rate (31). Both effects have been shown to persist through the cold pressor test. Furthermore, placebo effects, or changes in heart rate and subjective pain ratings due simply to the belief in heart rate change, do not account for reported subjective pain ratings or heart rate changes (24). In summary, biofeedback training has been shown to effectively produce an increased heart rate that persists through an acute stressor, and that cannot be accounted for by placebo. Whether the same is true for vagal tone is yet unclear.

The present study attempted to increase vagal tone by using biofeedback, while maintaining constant respiration. The durability of these increases was tested using the cold pressor test.

METHODS

Participants

Subjects were 13 male and 26 female undergraduates (mean age = 21 ± 2.3) from McGill University solicited using ads posted around campus. Strenuous physical activity and the use of over-the-counter medication, caffeine, and nicotine were prohibited for four hours prior to experimentation. Participants were compensated $10 for their time.

Apparatus

A Delta Biometrics Inc. vagal tone Monitor II was used to compute heart rate and vagal tone data. Subjects were comfortably seated approximately 2 feet from the feedback display in a padded chair. After alcohol swabbing, two hypoallergenic electrodes were placed on the lower rib cage, and a third ground electrode was placed on the ankle. The electrocardiogram obtained from these electrodes was amplified using a Grass polygraph and processed, in turn, by the Vagal Tone Monitor II, which quantifies beat-to-beat heart rate variability in the 0.15 to 0.40 Hz frequency band (i.e. respiratory sinus arrhythmia).

A 2-liter plastic container was filled with 2 trays of ice cubes and 1 liter of cold water at a temperature of 4 degrees Celsius for the cold pressor test. An electronic thermometer was used to insure a stable and adequate temperature. A wire mesh was placed diagonally in the tub to separate the ice from contact with the subject's hand.

In place of a metronome, a computer program simulating a 0.25 Hz sin wave was displayed approximately 2 feet from subjects on a laptop computer screen. Participants were initially trained to pace their respiration to an "X" moving along the sin wave.

Procedure

To assess the differential efficacy of vagal tone biofeedback and controlled breathing to increase RSA at rest and during the cold pressor test, subjects were randomly assigned to one of four experimental conditions. The vagal tone biofeedback condition (B), vagal tone biofeedback and paced breathing condition (PB+B), paced breathing only (PB), and sitting quietly (C). All groups practiced their respective manipulations for two 15-minute sessions.

Following informed consent and a brief introduction, all subjects sat quietly for 5 minutes in order to reach a baseline measure of vagal tone (see Table 1 for an outline of the experimental session). An initial 1 minute cold pressor test was administered to assess baseline vagal reactivity. Five minutes of quiet sitting followed to allow for vagal tone recovery. Subsequently, each experimental group had 2 sessions of 15 minutes to practice their respective manipulation. Between sessions, a 3 minute break was provided. After the second trial, all subjects were asked to retake the cold pressor test. Subjects sat quietly for another 5 minutes to allow for the final vagal tone reading. Subjects in the paced breathing group were asked to pace their breath with a 0.25 Hz moving sign wave.

Subjects in the biofeedback condition were asked to increase the displayed (vagal tone) readout. The true nature of the number was not revealed, although subjects were told that relaxation would help make the number rise. The biofeedback and paced-breathing group were both given the above instructions. This group was instructed to first feel comfortable breathing consistently with the computer program, and then move towards increasing the feedback display. Subjects in the control group were asked to sit quietly.

Table 1: Outline of Experimental Session

Procedure
Baseline- 5 min
Baseline cold pressor test- 60 sec
Rest 1- 5 min
Practice Session 1- 15 min
Break- 1 min
Practice Session 2- 15 min
Cold pressor test 2 - 60 sec
Rest 2 - 5 min

As suggested by Reyes et al. (26), subjects attempting vagal tone increases using biofeedback (groups B and PB+B) were informed that the feedback is normally quite variable, and that they should not be discouraged by seemingly incongruous feedback. Rather, they should focus on the general trend of the feedback readout.

Vagal tone measurements were taken in 5 second increments throughout the entire experimental session. Mean values of vagal tone were subsequently computed for each portion of the experimental procedure. Data files were transferred from the Vagal Tone Monitor II and analyzed using Microsoft Excel spreadsheets. Before debriefing, participants were asked to fill out a questionnaire describing their subjective experiences of each portion of the experiment, and their strategies for achieving the goal of their respective treatments. They were also asked to rate their subjective pain during the second cold pressor, and whether they found the second cold pressor to be more or less painful than the first.

Statistical Analysis

Repeated measures ANOVAs and t-tests were used to assess differences within and between experimental conditions. The time variable (within subject), treatment variable (between subject), and interactions were calculated using SYSTAT software.

Expected Results

It was hypothesized that the paced breathing plus biofeedback (PB+B) and biofeedback (B) conditions would produce the greatest increases in vagal tone. The control group (C) should show no increases, and the paced breathing group (PB) should show only moderate increases in vagal tone. Furthermore, although all the experimental groups should show less cardiac reactivity to the second cold pressor test, the paced breathing plus biofeedback (PB+B) group should demonstrate the greatest ability to mitigate the physiological arousal of this stressor.

RESULTS

The effects of the four conditions on seated, non-stress values of RSA, were examined using a 4 (group) X 3 (baseline, session 1, session 2) analysis of variance (ANOVA). This ANOVA revealed a significant interaction between condition and time ($F(6,70)=2.43$, $p=0.044$). All three treatment groups, (PB, B, PB+B) produced increases in vagal tone compared to controls. The paced breathing ($t(17) = 1.97$, $p= 0.034$) and paced breathing plus biofeedback ($t(18)=2.63$, $p = 0.032$) groups produced significant increases during the second practice session compared to controls. The biofeedback condition produced increases in vagal tone that approached significance ($t(18) =1.1$, $p=0.143$). T-tests revealed no significant differences between groups at baseline, but significant differences during the second practice session, indicating that the vagal increases were due to the experimental manipulations and not time (Figure 1 and 1b).

As expected, subjects in the four groups did not differ significantly in baseline vagal tone, or reactivity to the first cold pressor test. However, to ensure that all increases were not an artifact of slightly, though non-significant, increased baseline vagal tone, as was the case for the biofeedback plus paced breathing group, session 2 scores were subtracted from baseline scores and then re-subjected to analysis of variance and t-tests. No differences were observed in the adjusted results, supporting the reliability of the vagal increases.

To address the possibility that the experimental increases were of differing quality, first change scores were calculated. Differences between the first cold

Figure 1 Modulation of vagal tone during experimental session (± standard error)

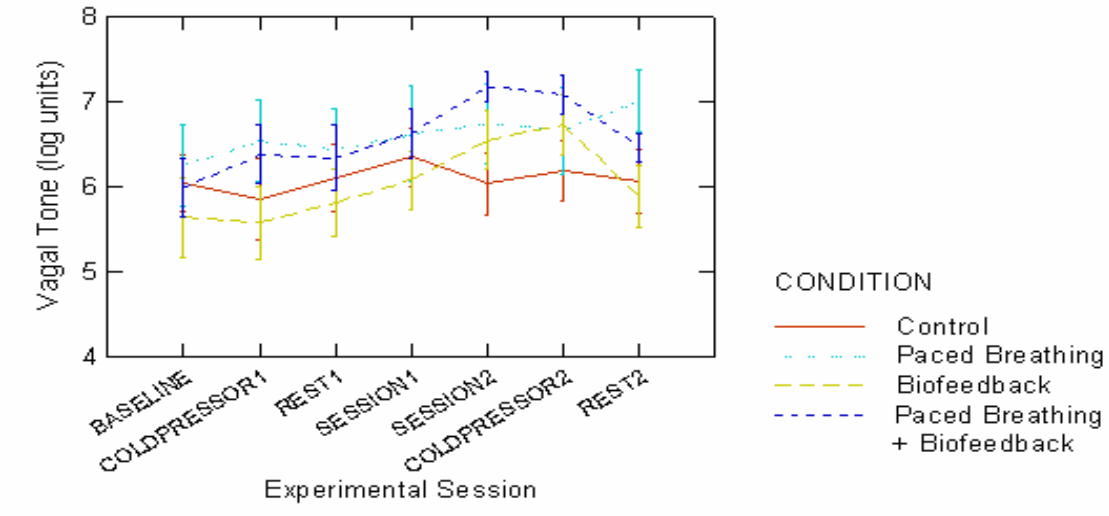
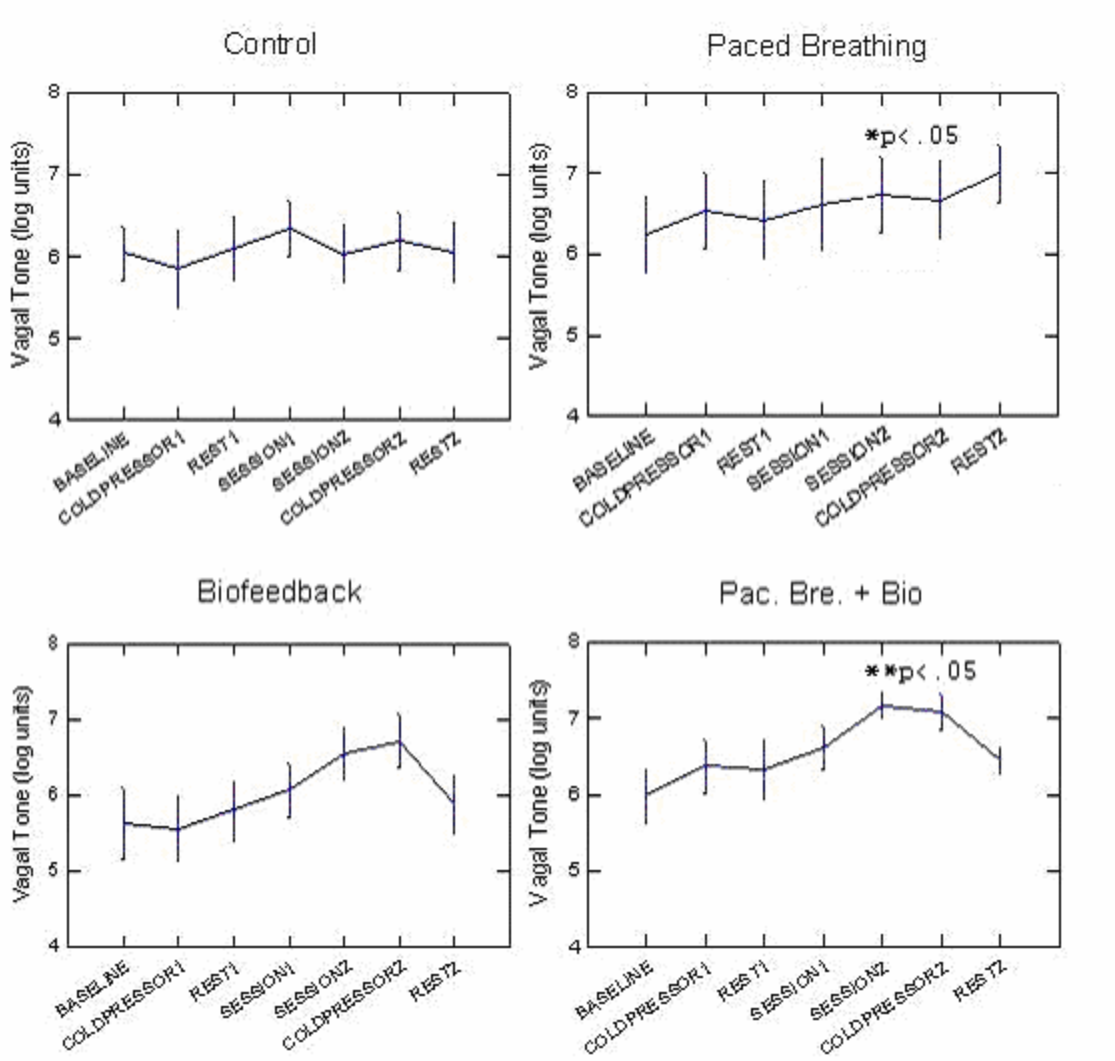


Figure 1b Modulation of vagal tone during experimental sessions. * (t (17) = 1.97, p= 0.034) ** (t (18)=2.63,p = 0.032) (± standard error)



pressor and baseline, and the second cold pressor test and session 2 were analyzed using repeated measures a 4 (group) X 2 (cold pressor) ANOVA. There were no significant interactions between condition and cold pressor reactivity (F (3,35)=0.878, p=0.46). The failure to find evidence to support differences in vagal quality is thought to reflect the insufficiency of the cold pressor to produce homogeneous physiological stress in subjects.

Similar analyses were performed on heart rate data. No significant interactions were found. However, as heart rate is not of primary concern to the present hypotheses, it will not be discussed further.

DISCUSSION

This investigation attempted to not only study whether biofeedback (B) and paced breathing (PB) could differentially increase vagal tone, but whether the combination (PB+B) could induce non-respiratory mediated increases in vagal tone. Furthermore, it was hoped that increases in vagal activity in the PB+B condition could better mitigate the stressful effects of the cold pressor test. Due to the failure of the cold pressor test to produce homogeneous physiological reactivity, only the prior question could be addressed.

Experimental Groups

Significant increases in vagal tone were observed in the paced breathing and paced breathing plus biofeedback conditions. These results replicate those of Reyes et al. (1992). As mentioned earlier, such increases in vagal tone are not surprising because increases in respiratory depth and frequency are known to boost vagal tone, and thus RSA.

Whether the significant increases in the paced breathing plus biofeedback group can be attributed to

the effects of the paced respiration, or to the production of non-respiratory mediated vagal tone is not overtly clear. It is possible that respiratory-based increases in vagal tone, as demonstrated by the paced breathing group, may have overshadowed any biofeedback effects. In this view, the observed vagal increases must be attributed to respiration. On the other hand, it is possible that the biofeedback protocol may have reinforced any respiratory effects, through its feedback conditioning, thereby producing more durable vagal tone. Subjecting these increases to a reliable stressor, such as timed serial subtraction or speech tasks, would have been helpful in determining the relative efficacies of respiration and biofeedback in the production of good quality vagal tone. It is unclear if such increases resemble those found by Sakakibara et al. (28) during autogenic relaxation.

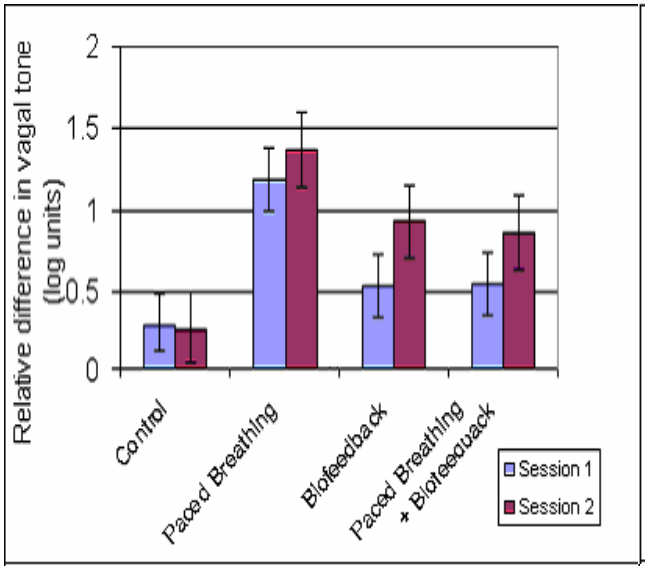
Although the biofeedback condition alone did not produce significant increase in vagal activity, increases were in the predicted direction. A number of factors may account for the demonstrated results.

Firstly, among all three experimental groups, subjects in the biofeedback condition had the most difficult task. They had to search for control strategies for the feedback by trial and error. It is not clear exactly how long such a search should take, although our data suggest that subjects improved from session 1 to session 2. In fact, as shown in Figures 2 and 3, this group produced the greatest-albeit non-significant-differences in vagal tone between the two practice sessions.

Secondly, in contrast to paced breathing, the practice session may not have been of sufficient duration to maximize the biofeedback effects. The data reveals this group to have the largest, though non-significant, difference between session 1 and session 2, indicating a longer learning period (Figure 2 and 3). Previous biofeedback research has used variable practice lengths and multiple practice sessions (19,26,28). Further investigation is needed to accurately gauge the optimal duration and frequency of practice sessions. Consideration, however, should be given to the clinical, financial, and logistical factors critical for the applied relevance of such research.

Thirdly, this group's feedback was of the visual type and highly variable, resulting in a lengthy period of adjustment. Subjects in the biofeedback group were required to keep their eyes open, and focus on a display. Such focused attention may have prevented them from fully relaxing, or concentrating on finding an adequate control strategy. The high variability of the output may have similarly hindered subjects' attempts at relaxation and control. It was felt that greater feedback data, in the form of 5 second intervals, was necessary to assist the subjects in the search for, and conditioning of, an accurate control strategy. More work must be done to investigate the trade-off between variability and feedback accuracy during biofeedback.

Similar issues may have also been present for subjects in the paced breathing plus biofeedback group. Their task, besides balancing the two required tasks, was to produce increases in vagal tone irrespective of respiratory manipulation. The experimental setup, however, may not have been ideal to induce such increases. Firstly, vagal increases due to the paced breathing may have become visible through the biofeedback display. Hence, subjects may have been reinforcing their respiratory maneuvers rather than producing increases from novel, non-respiratory



Figures 2 Relative differences in vagal tone between session 1 and session 2 (± standard error)

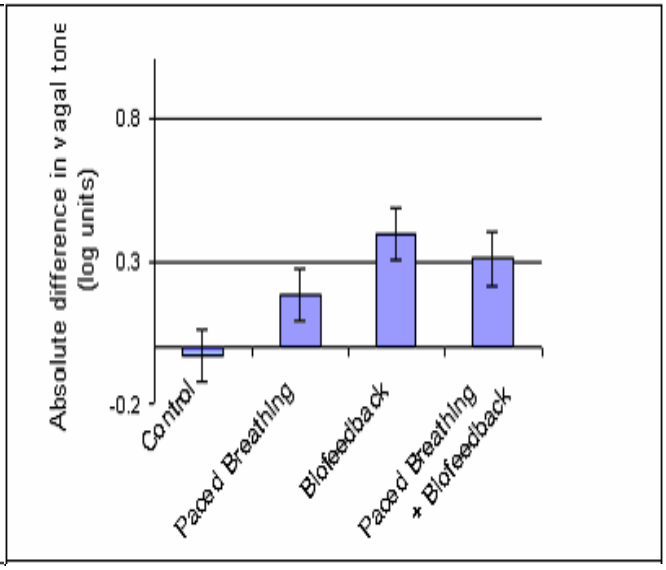


Figure 3 Absolute differences in vagal tone between session 1 and session 2 (± standard error)

related, control strategies. The use of a pneumogram would have been useful to ensure that subjects did not inadvertently alter their respiration in an attempt to increase the feedback display in all conditions.

Secondly, one cannot be sure of the extent and proportion of subjects' directed attention towards the biofeedback display. Although subjects were instructed to first master the paced breathing and then move on to the display, they may have found respiration a more "pleasant" and rewarding strategy, in which case their results would reflect the paced breathing group. Likewise, subjects may have focused only on the biofeedback task. A more harmonious integration of the paced breathing and biofeedback tasks, as well as the addition of a check to ensure compliance, would be of great value in future attempts to produce valid and reliable non-respiratory mediated increases in vagal tone. Assigning the tasks to two separate sensory modalities, by using auditory rather than visual feedback, for example, might ease the integration of the tasks.

Cold Pressor Test

As a result of the failure of the cold pressor test to evoke consistent decreases in vagal tone, and of the resulting fractioned subject group being too small for analysis, the quality of vagal increases could not be tested.

The subject population did not react homogenously to the cold pressor test. Only 17% (n=7) reacted to the cold pressor in the predicted fashion, with decreases in vagal tone to both cold pressors. However, 33% (n=13) of the subjects reacted with increases in vagal tone to both cold pressors. Although not verified, these increases may have been due to hemodynamic vasoconstriction, blood pressure increases, and subsequent baroreceptor medicated bradychardia. This hemodynamic response may have overshadowed any parasympathetic withdrawal, rendering the majority of the subjects inadequate for analysis. The remaining 50% of subjects (n=19) did not show consistent reactions in either direction to the cold pressor test. Their vagal activity was either neutralized by a possible hemodynamic response, or was altogether absent.

Additionally, the participants in the present study were young and healthy. Their cardiac and hemodynamic reactivity to the cold pressor stimulation exemplified a healthy "capacity to organize physiological resources" to respond to environmental demands (23). Experimentally speaking, what one gains in fine-tuned and efficient neural feedback by using such subjects, one loses in the homogeneity of stress reactivity. Likewise, it is unhealthy and comparatively older individuals to whom increased vagal tone is an asset. Thus, these results cannot necessarily be generalized to the population to which this research is applicable.

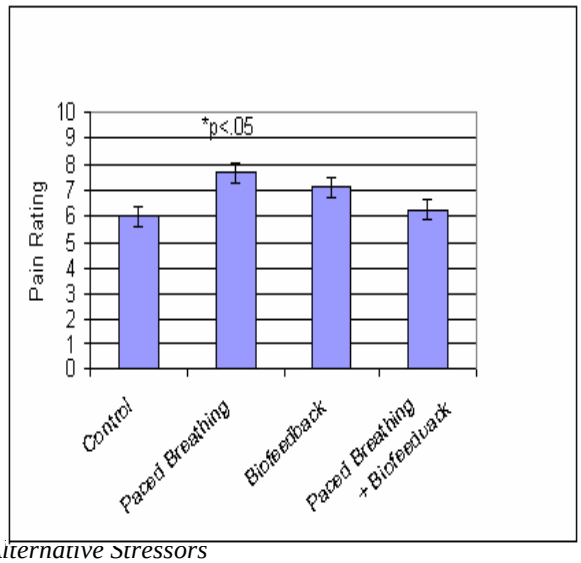
Although a commonly used stressor, both increases and decreases in vagal tone have been observed in response to the cold pressor test. Both water temperature and body part (i.e. forehead, foot, hand) that was subjected to stimulation, are important factors influencing cardiac and hemodynamic responses to stimulation (30). Since all participants in the present study were able to withstand the full 1 minute test, perhaps an increased stimulation period, or decreased water temperature, would have produced more pain and thus more consistent decreases in vagal tone. It is unclear, however, how older and unhealthy individuals might react to such stimulation.

The cold pressor test has been characterized as a passive physiological stressor (30). Such a classification has made it appealing as an experimental stressor, as ideally all subjects should respond alike without confounding psychological factors. Nonetheless, distinctly psychological coping skills and personality variables can shape the arousal profile of the cold pressor. For example, subjective pain during the cold pressor has been shown to correlate with increases in heart rate, and thus, decreases in vagal tone (Reeves, 1982). Subjects reporting pain would thus be expected to differentially experience greater decreases in vagal tone. Subjects in the paced breathing condition reported significantly greater ratings of pain than controls ($t(17) = 2.21$, $p = 0.042$). Contrary to past research, this group also exhibited the greatest overall increase in vagal tone, although their performance during the cold pressor tests remains uncertain. Regrettably, baseline pain ratings were not collected for the first cold pressor test (Figure 4).

Turner (30) suggests capitalizing on individual differences in physiological stress reactivity by forming reactor subgroups, such as "High and Low Heart Rate Reactors" or, in the case of this experiment: "Vascular and Cardiac Reactors." Although such an analysis was attempted on the present results, the resulting sample sizes were too small for analysis.

Interestingly, it may be possible to boost the proportion of subjects in either reactor subgroup. Groups such as gender and ethnicity have been shown to differ on these reactivity dimensions, thus providing an avenue by which to select a subject population predisposed to a particular reactor subtype. For example, relative to Caucasian Americans, African Americans tend to respond vascularly to stress induced increases in blood pressure (1). Future research using the cold pressor test should attempt to pre-select subjects prone to cardiac reactivity.

Figure 4 Subjective pain ratings of the second cold pressor test by group. * (t (17) =2.21, p = 0.042) (± standard error)



Alternative Stressors

Although the majority of contemporary stress-evoking events are not life threatening as they were during hunter-gatherer times, these challenges nevertheless elicit similar physiological arousal (30). Hence, psychological tasks involving active and continuous mental effort, such as serial subtraction, video games, and speech tasks, have all successfully been used to elicit reliable and stable physiological arousal.

Distinctly cognitive traits, such as locus of control and cynical mistrust, have also been shown to influence cardiac reactivity. Both an internal locus of control and high hostility, as measured by the Cook and Medley (8) hostility scale, can predispose an individual to greater cardiovascular reactivity (27,29). Similarly, Houston's (1992) model postulates a relationship between personality characteristics and cardiac reactivity via emotional and motivational arousal.

In order to mitigate the confounding effects of personality in research paradigms such as the present study, one must make a concerted effort to minimize any situational factors that could trigger a particular personality characteristic. Thus, for example, experimental instructions should be taped and carefully worded to avoid any desirability effects, and minimize any social and personality influences on reactivity. Such issues may have been present in this investigation.

CONCLUSION

Significant increases in vagal tone were produced using paced respiration, and the combination of paced respiration and biofeedback. However, due to the failure of the cold pressor test to induce consistent and homogeneous cardiac reactivity, and the possible confounding effects of vascular hemodynamic reactivity, this study was unable to address the durability of these increases.

The importance of vagal tone to cardiovascular health, however, is quite evident, especially in those individuals with cardiovascular disease. Evidence to support non-pharmacological methods to increase cardiac vagal activity; on the other hand, has gained only marginal application at the clinical level (19,28). Increasingly, however, a taxed and bewildered medical system is looking to non-conventional approaches to the management of chronic illnesses. Such approaches as Mindfulness Based Stress Reeducation (MBSR) have made great gains in substantiating behavioral and lifestyle modification as essential components of health. Dr. Jon Kabat-Zinn (16) has advocated responding to stress with mindfulness, or the complete acceptance of moment-to-moment events, thoughts, and emotions. Interestingly, MBSR involves much slower and deeper breathing. As previously discussed, ample evidence exists to support such breathing patterns as fostering increased cardiac parasympathetic activity. One need not dismiss slow breathing as efficacious simply because of its simplicity. Rather, the need for a convenient, easy to use, accessible, affordable, and efficacious strategy to increase RSA and vagal tone, makes respiration an attractive technique.

ACKNOWLEDGMENTS

The author would like to acknowledge the assistance of Michael Alper for kindly accommodating the lab space and time for this experiment as well as Max Lewkowski for his thorough and helpful revisions, advice, and support. In addition, the indispensable advice and support from Ms. Tone Liodden was of great impact to the thought and perseverance underlying the present paper. Furthermore, generous support from Dr. Blaine Ditto during all stages of this project was much appreciated. Both his and Dr. David Ostry's keen comments helped create a cogent and factually accurate paper.

REFERENCES

1. Anderson, N.B., McNeily, M., and Myers, H. Toward understanding race differences in autonomic reactivity, A proposed contextual model. In J.R. Turner, A. Sherwood, and K.C Light (Eds), Individual differences in cardiovascular response to stress. New York, Plenum, 1992.
2. American Heart Association. Heart Disease and Stroke Statistics -2003 Update. http,//www.americanheart.org/downloadable/ heart/ 1040391091015HDS_Stats_03.pdf 2003.
3. Ashby, W.R. An Introduction to Cybernetic. New York, John Wiley and Sons. Inc, 1963.
4. Benson, H., Beary, J.F, and Carol, M.P. The relaxation response. Psychiatry 37, 37-46; 1974.
5. Bentson, G.G, Cacioppo, J.T., and Quigley, K.S. Respiratory Sinus Arrhythmia, Autonomic origins, physiological

mechanism, and psychophysiological implications. Psychophysiology 30, 183-196; 1993.
6. Birbaumer, N., and Kimmel, H.D. Biofeedback and self-regulation. New York, Hillsdale, N.J. L. Erlbaum Associates, 1979.
7. Borovikova, L. V., Ivanova S., Zhang M., et al. Vagus nerve stimulation attenuates the systemic inflammatory response to endotoxin. Nature 405, 458-462; 2000.
8. Cookm W.W, and Medley, D.M. Proposed hostility and pharisaic-virtue scales for the MMPI. Journal of Applied Psychology 38, 414-418; 1954.
9. Eckberg. D.L. Parasympathetic cardiac control in human disease, A critical review of methods and results. American Journal of Physiology 239(5), H581-H593;1980.
10. Eckberg DL, Drabinsky M, Braunwald E. Defective cardiac parasympathetic control in patients with heart disease.. New England J Med. 285(16),877-883; 1971.
11. Eckholdt, K., Bodmann, K. H., Pfiefer, B., et al. Sinus Arrhythmia and heart rate in hypertonic disease. Advances in Cardiology 16, 366; 1976.
12. Fouad, F. M., Tanazi, R.C., Rerrario, et al. Assessment of parasympathetic control of heart rate by a non-invasive method. American Journal of Physiology 246, H838-H842; 1984.
13. Grossman, P., Stembler, G., and Meinhardt, E. Paced respiratory sinus arrhythmia as an index of cardiac parasympathetic tone during varying behavioral tasks. Psychophysiology, 27, 404-416; 1990.
14. Hayano, J., Yanada, A., Mukai, S., et al. Severity of arteriosclerosis correlates with the respiratory component of heart rate variability. American Heart Journal 121, 1070-1079; 1991.
15. Heart and Stroke Foundation of Canada. The Changing Face of Heart Disease and Stroke in Canada. Ottawa, Canada; 1999.
16. Kabbat-Zinn, J. Full Catastrophe Living. New York, Delta; 1990.
17. Katona, P.G, and Jib, R. Respiratory Sinus Arrhythmia, A noninvasive measure of parasympathetic cardiac control. Journal of Applied Physiology 39, 801-805; 1975.
18. Kleiger, R.E., Miller, P., Bigger, T., et al. Decreased Heart Rate Variability and Its Association with Increased Mortality after acute myocardial infraction. The American Journal of Cardiology 59, 256-261; 1987.
19. Lehrer, P., Smetankin, A., and Potapova, T. Respiratory Sinus Arrhythmia Biofeedback Therapy for Asthma, A Report of 20 Unmedicated Pediatric Cases Using the Smetankin Method. Applied Psychophysiology and Biofeedback 25(3), 193-200; 2000.
20. Lown, B., and Verrier, R.L. Neural activity and ventricular fibrillation. New England Journal of Medicine 364, 1165-1170; 1976.
21. O'Hair, D.E. Biofeedback, Review, History, and Application. http,//free.hostdepartment.com/d/deohair/psychoph.html 1988.
22. Porges, S.W. Respiratory Sinus Arrhythmia, Physiological basis, quantitative methods, and clinical implications. In P. Grossman, K.H. Jameson, & D. Vaitl. Cardiorespiratory and cardiosomatic psychophysiology. New York, Plenum Pub Corp,101-106; 1986.
23. Porges, S.W., and Vagal Tone, A physiologic marker of stress vulnerability. Pediatrics 90(3), 498-504; 1993.
24. Reeves, J.L., and Shapiro, D. Heart Rate biofeedback and cold pressor pain. Psychophysiology 19(4), 393-403; 1982.
25. Reeves, J.L., and Shapiro, D. Heart Rate Reactivity to the cold pressor stress following biofeedback training. Biofeedback and Self Regulation 8(1), 87-89; 1983.
26. Reyes Del Paso, G., Godoy, J., and Vila, J. Self-Regulation of Respiratory Sinus Arrhythmia. Biofeedback and Self Regulation 17(4), 261-275; 1992.
27. Rotter, J. Generalized expectancies of internal and external control of reinforcement. Psychological Monographs 80 (1), Whole No 609; 1966.
28. Sakakibara, M., Takeuchi, S, and Hayano, J. Effect of relaxation training on cardiac parasympathetic tone. Psychophysiology 31, 223-228; 1994.
29. Suarez, E.C, and Williams, R.B., Jr. Situational determinants of cardiovascular emotional reactivity in high and low hostile men. Psychosomatic Medicine 51, 414-418; 1989.
30. Turner, J.R. Cardiovascular Reactivity and Stress, Patterns of Physiological Response. New York, Plenum; 1994.
31. Victor, R., Mainardi, A., and Shapiro, D. Effects of Biofeedback and Voluntary Control Procedures on Heart Rate and Perception of Pain during the Cold Pressor Test. Psychosomatic Medicine 40(1), 216-225; 1978.

Dorey Glenn obtained his B.Sc. in Psychology from McGill University. He is currently an MA/MD student at Touro College and the Technion Institute of Technology in Haifa, Israel. He is interested in the use of mindfulness / awareness techniques, such as meditation, for health and well-being. His current career goals include practicing as a rural family physician or geriatrician.

ORIGINAL ARTICLE

Arterialised Capillary Blood Gases in Accident and Emergency Department Patients - a Reliable Alternative to Arterial Sampling?

Faheem Shakur *, Suzanne Mason, M.D.

OBJECTIVES: Many patients with respiratory complaints who present to the Accident & Emergency (A & E) department have an arterial blood gas analysis performed at some point. It is our belief that there is no difference between arterial and capillary blood gas values in patients presenting to the A & E department. It is also anticipated that body temperature and blood pressure may play a part, so these will also be reported and associations will be investigated. METHODS: Patients who require arterial blood gas analysis at any stage during their stay in the A & E department at the Northern General hospital of Sheffield are eligible for inclusion in the study. In total there were 32 patients. PROCEDURE:Transvasin cream was applied to the ear lobe to improve local blood flow by dilating the capillaries. When ten minutes have elapsed after the application of the Transvasin cream, a capillary sample is taken from the ear lobe by the researchers. CONCLUSION: From the t-tests conducted, no significant difference was seen between the arterial and capillary blood gas samples for the parameters pO_2 and O_2 saturation. However, for pCO_2, pH and [HCO_3] there were significant differences observed. This result seems to disagree with the findings of most other studies that have so far shown stronger correlations generally for pH, pCO_2 and bicarbonate, than for oxygen measuring parameters.

BACKGROUND

Many patients with respiratory complaints who present to the department have an arterial blood gas analysis performed at some point. Normal blood gas values are as follows: pH 7.34-7.45, pO_2 80-100 mmHg and pCO_2 of 35-45mmHg.The idea of an arterial blood gas is to assess the oxygenation of arterial blood and the blood's acid-base balance (1). This test is helpful in determining the root cause of their presenting complaint and it assists in defining and monitoring the necessary treatment. It is a painful procedure that involves taking a small blood sample from the radial artery and complications, such as formation of a hematoma and development of infection, are not uncommon (2). Patient preparation includes assessment of peripheral circulation on both sides. The artery of choice is the radial artery, which is usually easily palpated at the flexor surface of the arm. If the radial artery is unsuitable as a puncture site, the brachial artery is the second choice, followed by the femoral artery. It is evident that such a procedure can at times be challenging to perform and is generally only performed by medical personnel. A simpler alternative for blood gas analyses has therefore been sought for many years.

It is our belief that there is no difference between arterial and capillary blood gas values in patients presenting to the A & E department. It is also anticipated that body temperature and blood pressure may play a part, so these will also be reported and associations will be investigated.

Several studies, performed as early as the beginning of the 1970s, investigated whether the blood gas results

* To whom correspondence should be addressed: Faheem Shakur, University of Sheffield, School of Medicine, Beech Hill Road, SHEFFIELD S10 2RX, UK.
E-mail address: faheemshakur@hotmail.com

from arterialised capillary vessels of the ear lobe would correlate with arterial ones and whether the former technique could replace the latter (3-6). This would have clear benefits as capillary blood sampling is a much easier technique, which can be performed by medical technicians, causing less discomfort to the patient with adverse events being very rare (2). Despite several reports that have supported the use of capillary sampling, arterial blood gases remain to date the 'gold standard'.

Most of the studies performed so far have come from specialised units, such as paediatric wards (7,8), or concentrated in the management of specific conditions such as diabetic ketoacidosis (9). Some of these studies have provided certain evidence to support the use of capillary blood sampling in certain cases. There have been as yet no studies that have investigated the use of the technique in patients with a wide variety of medical conditions, such as those that occur in an A & E department. In such a busy setting the need for arterial blood sampling can at times cause delays in the diagnosis and management of the patient. The question that arises is whether an easier and quicker technique such as capillary blood sampling could give results that are comparable to those coming from arterial sampling, and therefore serve as a reliable and practical alternative.

METHODS

Inclusion and exclusion criteria

Patients who require arterial blood gas analysis at any stage during their stay in the Accident and Emergency department at the Northern General hospital of Sheffield, are eligible for inclusion in the study. Patients who require blood gas analysis often experience respiratory distress and on occasions their consciousness may be impaired. Such patients cannot voluntarily take part in the study and were therefore excluded. All other patients were informed about the study and must have been able to understand the procedure before it was undertaken. Consent must be asked for and the necessary forms signed by the patient, and the person conducting the procedure. Participation in the study did not affect the patient's management, and our findings are kept confidential. If for any reason adequate blood cannot be taken from the ear lobe, consent forms and other necessary documents are retained. The research project received ethical approval by the North Sheffield Local Research Ethics Committee.

Procedure

Transvasin cream, commonly used for muscular pain, was applied to the ear lobe to improve local blood flow by dilating the capillaries. Pulse, blood pressure and temperature are recorded while an arterial sample is obtained in the usual manner, via the radial artery. Both the radial and the ulnar arteries are compressed at a level approximately 1cm proximal to the wrist joint while the patient makes a tight fist for approximately 5 seconds. The patient is then instructed to open the fist in a relaxed fashion. About 2 ml of arterial blood is then collected from the radial artery, a procedure carried out by the attending A & E doctor. This sample was then analysed as soon as possible.

When ten minutes had elapsed after the application of the Transvasin cream, a capillary sample was taken from the ear lobe by the researchers. A puncture was made using a unilet lancet, a drop of blood was allowed to form and was then allowed to suspend from the earlobe. A sample was collected in a 175 mm capillary tube from the centre of the drop of blood. Analysis of the sample was undertaken within no more than five minutes from collection, using the Ciba-Corning 865 ABG analyse. The researchers collecting the data were medical students working with the cooperation of A & E doctors who were doing the routine collection of the arterial blood gases.

Methods of data analysis

Once both samples had been analysed, the pH, pO_2, pCO_2, $[HCO_3]$ and oxygen saturation levels were recorded for further statistical tests. The strength of relationship between the arterial and capillary blood gas results was analysed using scatter plots, correlation coefficients and t-tests. A two-tailed t-test was used for non-independent samples, at a significance level of $p=0.05$. This test assumes the data is distributed normally and so plots were carried out to prove this assumption. (appendix 1)

For each set of pH, pCO_2, pO_2, $[HCO_3]$ and O_2 sat values, the data from the capillary samples was plotted against that from the arterial samples.

The correlation coefficient is a measure of how well trends in the predicted values follow trends in the actual values.

RESULTS

Data set

32 patients were included in the study. Of these, 19 were males ranging in age from 42 to 86, and 13 were females aged between 19 and 92. The procedure was successful, and a capillary sample obtained, in 26 of the 32 individuals.

DISCUSSION

From the t-tests conducted no significant difference was seen between the arterial and capillary blood gas

Table 1. T-test results of measured blood parameters

T-test Results	
Mean pH (Arterial)	**7.41**
Mean pH (Capillary)	**7.39**
t-test value	**5.54**
Critical (p) value	**2.074**
Mean pCO_2 (Arterial)	**4.88**
Mean pCO_2 (Capillary)	**3.84**
t-test value	**5.43**
Critical (p) value	**2.074**
Mean $p0_2$ (Arterial)	**12.5**
Mean $p0_2$ (Capillary)	**13.16**
t-test value	**-0.49441**
Critical (p) value	**2.074**
Mean $HC0_3$ (Arterial)	**23.53**
Mean $HC0_3$ (Capillary)	**20.67**
t-test value	**2.66**
Critical (p) value	**2.08**
Mean 0_2 Sat (Arterial)	**96.99**
Mean 0_2 Sat (Capillary)	**96.6**
t-test value	**1**
Critical (p) value	**2.571**

samples for the parameters pO_2 and O_2 saturation. However, for pCO_2, pH and $[HCO_3]$, there were significant differences observed. This seems to disagree with the findings of most other studies that have so far shown stronger correlations generally for pH, pCO_2 and bicarbonate, than for oxygen measuring parameters (7-9,11-16).

Any problems faced in the preparation and the data collection periods could have affected the overall results. With more training, the technique of taking a capillary sample would have made the study more successful. Approximately 20% of the patients willing to take part in the study were not actually able to participate as an insufficient amount of blood could be taken from the ear for a proper reading. A possible way to get around this problem would be to use capillary samples from the fingertip, where more blood could possibly be obtained. This technique has been successful in some studies (9), though it is associated with more pain, removing one of the major advantages of capillary sampling. On occasions researchers may have failed to stab the correct part of the ear - the inferolateral aspect of the pinna (10). Also, as the capillary sample was small, if any air were to contaminate it, there would be a large difference noticed. However, since the arterial sample is larger, a small amount of air would not noticeably affect the overall results. Air contamination causes the pO_2 to be raised.

It was extremely difficult to conduct the capillary sampling at the same time as the arterial sampling. A time delay occurred, due to the 10 minute wait for the Tranvasin cream to take effect. Taking blood from the fingertip would help reduce this time delay (9). Due to the time delay between the two samples, the patient could have been administered oxygen during the interval and clearly, this would have altered the results. Studies that conducted the two methods simultaneously had the best results (2). Due to the nature of the setting, if oxygen needed to be administered to a patient, the treating doctors of the A & E department could not wait for the researchers to conduct the capillary sampling. Occasionally, when a capillary sample was taken there was a queue for the analysing machine. Putting the sample in ice to prevent to degeneration of white cells may have helped in that circumstance. Other technical problems were also encountered such as machine calibration faults.

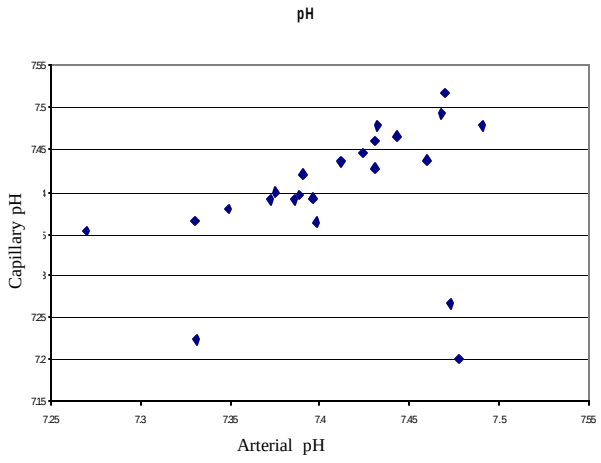


Figure 1. Scatter plots and correlations of Arterial and Capillary parameters.The scatter plot for pH shows an apparent positive correlation between the arterial and capillary samples, consistent with the two methods giving similar results.

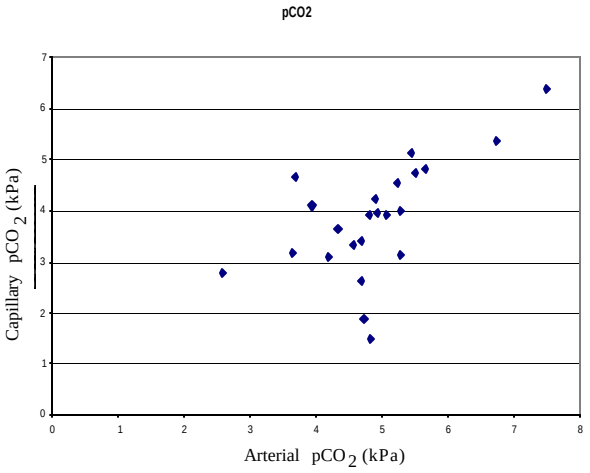


Figure 2. The scatter plot for pCO_2 seems to show a positive correlation between the arterial and capillary samples, with a definite clustering around 4.5/5 kPa, consistent with the two methods giving similar results.

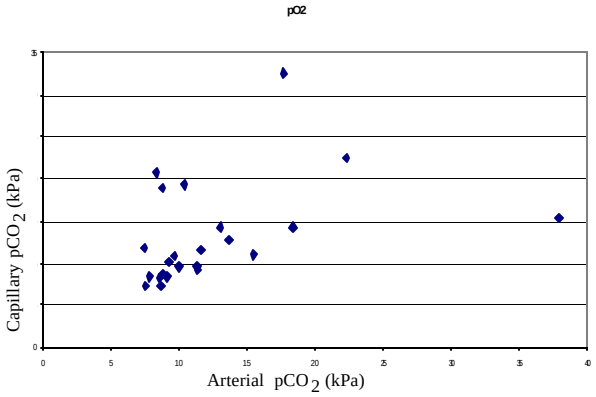


Figure3. The scatter plot for pO2 shows an apparent positive correlation between the arterial and capillary samples with a large clustering around 10kPa and a few points, which are very scattered, consistent with the two methods giving similar results.

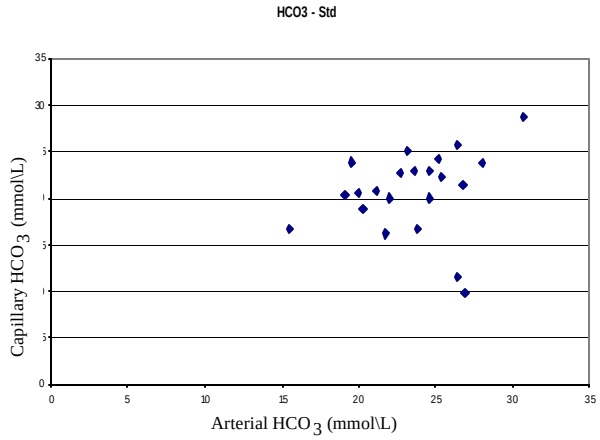


Figure 4. The scatter plot for HCO_3 shows a fairly positive correlation between the arterial and capillary samples, consistent with the two methods giving similar results

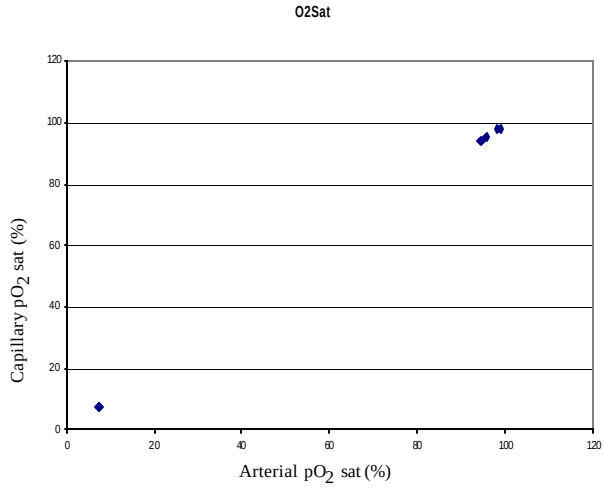


Figure 5. As there were very few samples taken for the O2 saturation it is hard to tell how the strong the correlation is between the methods however we are able to see that there is still a positive correlation.

Greater accuracy and a reduction in bias could have been achieved had the sample size been larger. Due to time limitations the study had to be restricted to a small population size, which compromises the significance of the results. From data that was previously collected it has been calculated that a sample size of 165 patients will be required to demonstrate a difference in pH of 0.03, calculated with 80% power at p=0.05 level. Similarly, a sample size of 232 patients would be needed to detect a difference in pCO_2 of 0.05 mmHg at the same power and significance level. Finally, a sample size of 332 patients would be required to detect a difference in pO_2 of 0.05 mmHg at the above levels of power and significance. We would therefore recommend that the data collection be continued up to a sample size of approximately 332 patients. If the Senior House Officers were collecting the capillary sample every time they took an arterial one, the desired sample size could probably be reached in a much smaller time period. Time delays between the two samplings could also be avoided to a great extent.

Our study was also too small to show any correlation of the main variables measured with data such as patient sex, age, pulse, temperature, blood pressure, respiratory rate, oxygen therapy, sample volume and quality. From the above list, blood pressure, as an indication of the hemodynamic condition of the patient, seems to be of great significance (7,8). Further studies could also try to correlate the value of inter-changeability between the two techniques for various specific conditions. These studies may also attempt varying the severity of these conditions and varying the patient age groups.

The importance of identification of patients definitely requiring admission at the point of triage has been highlighted at the recent government paper "Reforming Emergency Care" (10). This strategy would serve the concept of "streaming," allowing patients to be admitted to a hospital bed in the most efficient manner. Other than routine observations and a brief clinical assessment, few tools are available to the triage nurse at present to facilitate this function. It may be that capillary blood gases could aid this process. In conclusion the authors would feel satisfied with capillary blood gas analyses if they were patients in the A & E department.

REFERENCES

1. Driscoll P. A Simple Guide to Blood Gas Analysis. BMJ Publishing Group. 1997
2. Dar K, Williams T, Aitken R, et al. Arterial Versus Capillary Sampling for Analysing Blood Gas Pressures. British Medical Journal. 1995; 310: 24-5
3. Sadove MS, Thomason RD, Jobgen E. Capillary Versus Arterial Blood Gases. Anesthesia and Analgesia. 1973; 52: 724-7
4. Osugi T, Shibata M. Clinical Application of Blood Gas Analysis

of Arterialized Capillary Vessels of The Ear Lobe. Iryo. 1971; 25: 660-3
5. Keller R, Herzog H. Comparison of Arterial and Capillary Blood Gas Analysis In Subjects with Clinical Evidence of Shock. Klinische Wochenschrift. 1971; 49: 54
6. Nonaka H, Tsurutani H, Okubo Y, Nagano H. Analysis of Arterialized Capillary Blood By Ear Lobe Method. Iryo. 1971; 25: 29-32
7. Harrison M, Lynch J, Dean JM, Witte M. Comparison of Simultaneously Obtained Arterial and Capillary Blood Gases In Paediatric Intensive Care Unit Patients. Critical Care Medicine. 1997; 25
8. Escalante-Kanashiro R, Tantalean-Da-Fieno J. Capillary Blood Gases in a Paediatric Intensive Care Unit. Critical Care Medicine. 2000; 28
9. Hale P, Nattrass M. A Comparison of Arterial and Non-arterialized Capillary Blood Gases in Diabetic Ketoacidosis. Diabetic Medicine. 1998; 5: 76-78
10. Pitkin AD, Roberts CM, Wedzicha JA. Arterialised ear lobe blood gas analysis: an underused technique. Thorax. 1994; 49: 364-6
11. Sauty A, Uldry C, Debetaz L-F, Leuenberger P, Fitting J-W. Differences in PO2 and PCO2 Between Arterial and Arterialized Earlobe Samples. European Respiratory Journal. 1996; 9: 186-9
12. Saili A, Dutta AK, Sarna MS. Reliability of Capillary Blood Gas Estimation in Neonates. Indian Paediatrics. 1992; 29: 567-70
13. Mathur US, Manchanda A, Singh V, Rishi JP, Athaiya V. Comparative Study of Capillary and Arterial, Blood Gas Values in Plastic and Glass Syringes. Indian Journal of Chest Diseases and Allied Sciences. 1989; 31: 247-50
14. Ridley J, Nadeem S, Sridhar M. Capillary Blood Gas Analysis for Long Term Oxygen Therapy Assessment. Clinical Medicine. 2001; 1: 326
15. Hughes JM. Blood Gas Estimations from Arterialized Capillary Blood Versus Arterial Puncture. European Respiratory Journal. 1996; 9: 184-5
16. Pandit JJ. Sampling for Analysing Blood Gas Pressures. Arterial Samples Are The Best. British Medical Journal. 1995; 310: 1071

Faheem Shakur is a third year medical student at University of Sheffield, UK.
Dr Suzanne Mason is a senior lecturer at SCHARR (School of Health and Related Research).

ORIGINAL ARTICLE

Annotating the Mycobacterium avium Genome: A Project in Bioinformatics

David D. Shersher[1†], Maksim Kirtsman[1†], Mikael Katz-Lavigne[1†], Marcel A. Behr, MD MSc[2], Makeda Semret, MD MSc[2†*]

ABSTRACT Bioinformatic tools facilitate efficient processing and formatting of experimental data and are becoming essential to research in the biological sciences. Whole genome sequencing projects, combined with DNA microarray technology, have allowed genomic comparisons between and within species of microorganisms. The genome of *Mycobacterium avium subsp. avium* (MAA) has been sequenced by The Institute for Genomic Research (TIGR), but a final and annotated version has not yet been made available. The goal of this project was to annotate the sequence of MAA as a foundation for microarray-based genomic comparisons. We used software to identify and predict open reading frames (ORFs) present in this organism. The ORFs were then compared to those catalogued in two large, online genetic databases for other microorganisms and matched to homologous sequences, allowing the determination of putative functions for each predicted gene. The genome of MAA was determined to contain 4480 genes, the majority of which are homologous to genes found in other Mycobacterial species.

Keywords: *Mycobacterium avium* complex, annotation, genomics, BLAST

INTRODUCTION

The combined availabilities of whole genome sequence information and new bioinformatic tools have fueled genomic research, allowing extensive comparisons between closely related species of microorganisms. Sequencing of the *Mycobacterium tuberculosis* genome in 1998 (1) laid the foundation for detailed study of the members of the *Mycobacterium tuberculosis* complex, which in turn has guided the establishment of evolutionary scenarios for this group of microorganisms as well as permitted inferences about their biology (2-4).

Since then, over 140 bacterial genomes, including several members of the M. tuberculosis complex, have been sequenced and genomic information has been made publicly available on the internet. In contrast, little genomic information is available for the *Mycobacterium avium* complex (MAC), a group of environmental organisms with tremendous pathogenic potential. The MAC is composed of two species, M. avium and M. intracellulare. The former is further subdivided into three subspecies, M. avium subsp. avium (MAA), M. avium subsp. paratuberculosis (MAP) and M. avium subsp. silvaticum (MAS). Defined as a complex based on their genetic similarities, these subspecies are phenotypically quite distinct. MAP is an obligate intracellular pathogen that is the cause of a chronic enteritis known as Johne's disease in cattle; it has also been suggested by some to be the cause of an inflammatory bowel disease in humans called Crohn's disease (5). MAA is predominantly an environmental organism which

*To whom correspondence should be addressed: Makeda Semret MD MSc FRCP(C), McGill University Health Centre. Research Institute RS1-105, 1650 Cedar Avenue , Montreal Quebec H3A 1G4 Canada, tel: (514) 934 1934 x 44621; fax: (514) 943 8261; e-mail:makeda.semret@mail.mcgill.ca .

1 Department of Microbiology and Immunology, Faculty of Medicine, McGill University, Montreal, Quebec, Canada.

2 Department of Infectious Diseases and Medical Microbiology, McGill University Health Centre

† These authors contributed equally to the manuscript.

causes a tuberculosis-like disease in birds, and which can also elicit a disseminated disease in immuno-compromised individuals. *M. avium subsp. silvaticum* also affects birds, although little is known about its true distribution. Moreover, members of the MAC present widely different laboratory characteristics. Their divergent patterns of disease manifestation and laboratory phenotypes are poorly understood but are likely to result from subtle variations within the organisms' respective genomes.

Genome sequencing projects for two of the members of the MAC (MAA and MAP) have been undertaken. Sequence information on MAA strain 104, a clinical isolate from an AIDS patient, has been released on the internet (TIGR, http://www.tigr.org), but a fully edited and annotated version of the genome is not yet available. The goal of this project was to annotate the complete genome sequence of MAA 104 in order to identify putative genes of this organism that will in turn serve as a template for DNA-microarray based genomic studies of the MAC. We describe herein the annotation of the 5.5 Mbp genome of MAA, which was accomplished through the use of various automation scripts.

METHODS

Identification of open reading frames

The 5.5 Mbp genomic sequence of MAA strain 104 was obtained from The Institute for Genomic Research (TIGR, http://www.tigr.org). Annotation was performed using Artemis™, a genomic visualization software package developed by The Sanger Institute (http://www.sanger.ac.uk/software). This tool allows the identification of open reading frames (ORFs) based on recognition of start and stop codons, and the in-silico translation of these nucleotide sequences into their respective amino acids. The minimum size of the ORFs identified was set at 100 bp. The Artemis™ version used (version 4) had limitations with respect to the size of the sequence that could be viewed at one time. For this reason, the 5.5 Mbp genome was split into 61 contiguous sequences (contigs) of 92,000 bp each. The contigs contained 500 bp overlaps at both ends in order to facilitate their later merging as well as to ensure that open reading frames spanning two adjacent contigs are appropriately identified.

The nucleotide sequences of each of the contigs was entered into Artemis™; once putative ORFs were assigned, the amino acid sequence of each ORF in the 61 contigs was exported into text files to permit alignment with sequences already deposited in public internet databases.

Alignment searches for predicted open reading frames

Alignment of each ORF sequence to existing databases was done using the Basic Local Alignment Search Tool (BLAST) (6). The BLAST algorithm allows comparisons of nucleotide or amino acid sequences for homology to previously annotated sequences or proteins that are deposited in public databases. For each predicted ORF within the MAA genome, Artemis™ provides an amino acid sequence that is then compared (using the blastp protocol) against amino acid sequences from other genomes. If those sequences had previously been assigned a probable gene function, by extension, we assigned the function of the most similar amino acid sequence to the corresponding MAA ORF.

The two major databases used were the *Mycobacterium tuberculosis* genome database (Release 5) from the Pasteur Institute (http://genolist.pasteur.fr/TubercuList/), and the Non-Redundant (NR) database from the National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/). Initially, the databases were accessed through an online BLASTing program called 'NetBlast'. This program takes local query sequences contained in text files and aligns them individually against the databases on their respective servers. The BLAST queries are queued based on the volume of searches occurring simultaneously, which causes delays proportional to the total number of queries. The solution to the query-delay problem was to use a local version of the BLASTing software called 'blastall' (ftp://ftp.ncbi.nih.gov/blast/) and to download and locally store the two databases.

Query batching

In order to query all of our MAA genome files in an easy, automated manner, text-based batch files were created. Both these files and the 'blastall' program are MSDOS™-based. The batch files were simple and contained the program execution command specifying the name and certain details of each contig file to be processed. Each BLAST result is reported with an Expectation (E) value, representing the number of hits with equal or better alignment scores than what one can expect to find by chance alone when searching a database of a certain size. Because the NR database is much larger than R5, we made the simplifying assumption that an E value of 0.01 in the former is more or less equivalent to an E value of 0.1 in the latter, and arbitrarily assigned these E values as our initial cutoffs.

Creating Microsoft Excel™ databases

In order to extract the useful information from the BLAST results and display it in a user-friendly format,

a Microsoft Windows™-based automation software was used. We custom-wrote a script to locate the desired information within the BLAST result file and transfer it to a Microsoft Excel™ spreadsheet. Furthermore, the E values obtained from homology searches against both databases were compared, and only the best hit (with the lowest E value) was imported into the Microsoft Excel™ spreadsheet. When Expectation values were similar, precedence was given to the TubercuList database as there is significant homology between M. avium and M. tuberculosis. If a putative ORF in MAA did not show a significant degree of homology to proteins in the M. tuberculosis database, the results from NR, if any, were retained. The script evolved numerous times during the process to improve on accuracy and efficiency. The fifth revision of the script had a proofreading function that could detect errors in the data transfers and correct them.

For each contig, the Microsoft Excel™ file generated was "cleaned" using another automation script that removed the empty ORFs with no hits in either of the two databases. For further processing, the Microsoft Excel™ files had to be reformatted into GFF ("Gene-Finding Format") files to permit future visualization in Artemis™.

Removal of overlapping ORFs in Artemis™

When the "cleaned" GFF files were imported into Artemis™, a graphical overview of the nucleotide and amino acid sequence with ORFs marked in the 6 possible reading frames was produced. The entire genome thus represented was visually scanned, specifically searching for overlapping ORFs. Once an overlap was detected, the longer ORF with the lower E value or the ORF with significant homology to a protein in R5, was retained while the other was manually removed. For example, if an ORF annotated as Equine Herpes Virus Protein with an E value of 0.01 overlapped another ORF whose best hit was with a M. tuberculosis gene that had an E value of 0.1, the *M. tuberculosis* homologue was retained.

Merging the contigs

The next step in the annotation was the merging of the GFF files based on the intentional overlaps left when the MAA genome was first split into contigs. This was done in Microsoft Excel™ by opening the manually edited GFF files and checking for identical ORFs found in the extremes of two adjacent contigs. The whole genome was thus reconstructed from the overlapping contigs.

Functional classification of the annotated genes

All putative gene products were subdivided into 10 functional categories of proteins based on the classification developed by the Pasteur Institute for the genomes of *Mycobacterium tuberculosis* and *Mycobacterium leprae*. The functional classification data was imported into Artemis™ by editing a GenBank (GBK) file of the annotation. A color-coding scheme was then used to separate the ORFs based on these different functional gene categories.

RESULTS

Annotation of the 5.5 Mbp chromosomal sequence of MAA strain 104 took several months. The initial 61 contigs were processed through the Artemis™ software package, and a total of 29,000 possible ORFs were identified. The contigs were subsequently batch-BLASTed, which generated two sets of files with the BLAST results: one set that showed homology results against the NCBI database and the other against the M. tuberculosis database. By applying the cut-offs for E values described above, we were able to decrease the number of ORFs to about 10,000.

After the final steps of removing overlapping ORFs and merging the 61 contigs, annotation was complete. A total of 4,480 putative ORFs were identified; 4,095 of these ORFs (91%) matched genes in the M. tuberculosis database (R5), while the rest matched to other organisms. Figure 1 shows an example of the

Functional Category*	ORFs in MAA 104 (% of total)
0 virulence, detoxification, adaptation	148 (3.3%)
1 lipid metabolism	436 (9.7%)
2 information pathways	222 (4.9%)
3 cell wall and cell processes	662 (14.8%)
5 insertion sequences and phages	155 (3.5%)
6 PE/PPE	53 (1.2%)
7 intermediary metabolism and respiration	1139 (25.4%)
8 unknown	93 (2.1%)
9 regulatory proteins	265 (5.9%)
10 conserved hypotheticals	862 (19.2%)
Unclassified (MTB orthologs)	60 (1.3%)
Unclassified (non MTB orthologs)	385 (8.6%)
Total # of genes annotated	4480

*functional categories are the same as those for the classification of *M. tuberculosis* (http://genolist.pasteur.fr/TubercuList/)

Table 1. Summary of the annotated M. avium subsp. avium genome, with functional classification of the predicted genes as per the classification for M. tuberculosis (http://genolist.pasteur.fr/TubercuList/). ORF = open reading frame, MAA = M. avium subsp. avium, MTB = M. tuberculosis. PE/PPE refers to a family of genes commonly found in the M. tuberculosis genome defined by the amino acid sequence motifs Pro-Glu (PE) and Pro-Pro-Glu (PPE). These motifs are located at the highly conserved N-terminal domains

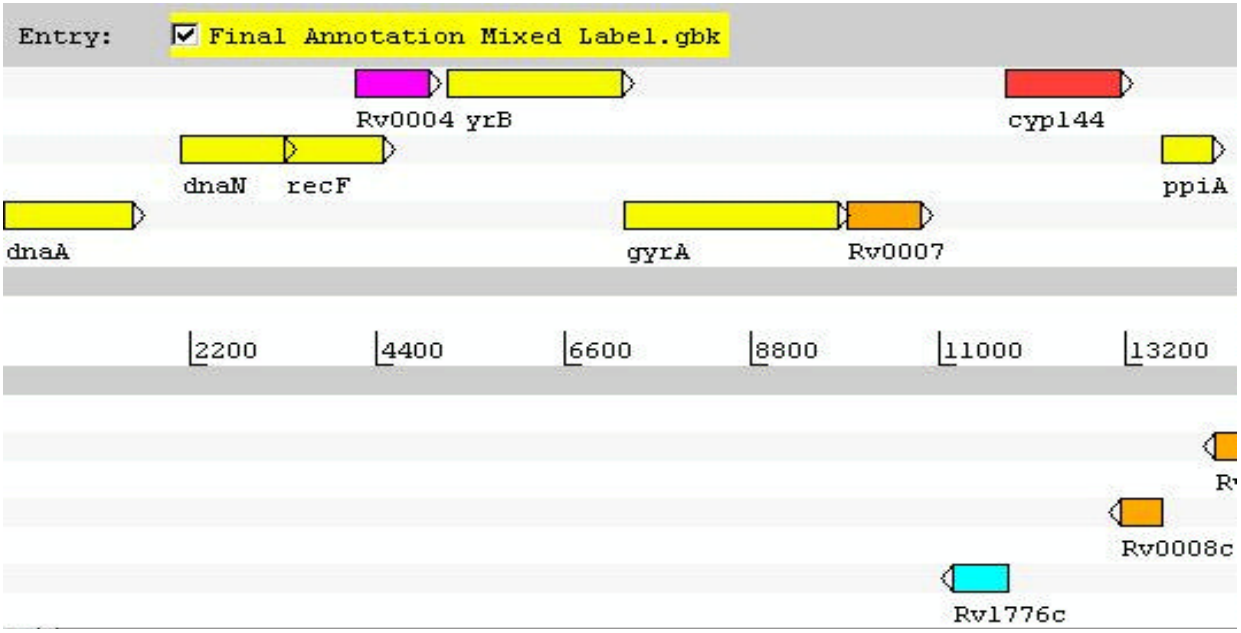


Figure 1. Screenshot of the final visual annotation as seen through Artemis. Colored ORFs represent genes in M. avium subsp. avium and are color-coded based on functional categories of genetic function. As there are 6 ways in which DNA sequence can encode amino acids (three frames forward and three frames in reverse), these colored arrows representing putative genes are found in any of these 6 reading frames. The direction of the colored arrow representing the ORF indicates the coding sense.

appearance of the ORFs in Artemis™. In this final annotation, there are few overlapping regions between the 4480 putative genes. Each of the ORFs is also color-coded based on its functional classification as defined by TubercuList (http://genolist.pasteur.fr/TubercuList /help/classif-search.html#Codes). Table 1 summarizes the number of ORFs per functional category and their percentage within the genome.

DISCUSSION

The *Mycobacterium avium subsp. avium* annotation project evolved from a need to cross-reference a previous annotation, which was performed in-house in the year 2000 but was based on a fragmented (non-sequential contigs) genomic sequence of MAA strain 104. This first annotation identified 5,574 genes that served as a template for the design of oligonucleotides used in the assembly of a DNA microarray. The array has since been used to perform extensive genomic comparisons between the members of the MAC and has been validated for transcriptome analysis for this species. However, the fragmented sequence released earlier most likely contained redundancy and errors. The release of an edited and circularized version of the MAA 104 sequence in June 2002 provided the opportunity to annotate the genome presenting the genes in their correct order, starting with dnaA in the origin of replication and ending with the gene rpmH.

The 5.5 Mbp genome of MAA 104 thus annotated contains 4,480 ORFs, for a gene density of 1,222 bp/gene. In contrast, the 4.4 Mbp genome of M. tuberculosis has 3,924 ORFs for a gene density of 1,114 bp/gene (1). Considering that there is a significant degree of genetic homology between Mycobacterial species, our annotation of the MAA genome may have slightly underestimated the total number of ORFs. Previous work has established that there is 97% homology between MAA and MAP over several large regions (7,8). In order to gauge the accuracy of our annotation, we compared one region of our annotated genome (the region surrounding the origin of replication, oriC) to the corresponding region in the MAP (8) and the M. tuberculosis genomes. We note over the oriC region that the number of genes and their order were similar among these three genomes (Figure 2), providing reassuring validation to our annotation.

The annotation process led to the identification of beneficial informatic tools. For instance, the version of the Artemis™ software used in this project (version 4) had inherent limitations in the size of the sequence that could be viewed at once. This was resolved by artificially splitting the genome into 61 files. The most recent release of Artemis™ (version 5) is able to display the entire 5.5Mbp genome without division into multiple contigs. As another example, some of the repetitive steps involved in annotation can be expedited by custom-written automation scripts, and/or with the use of Visual Basic™ macros for performing some of

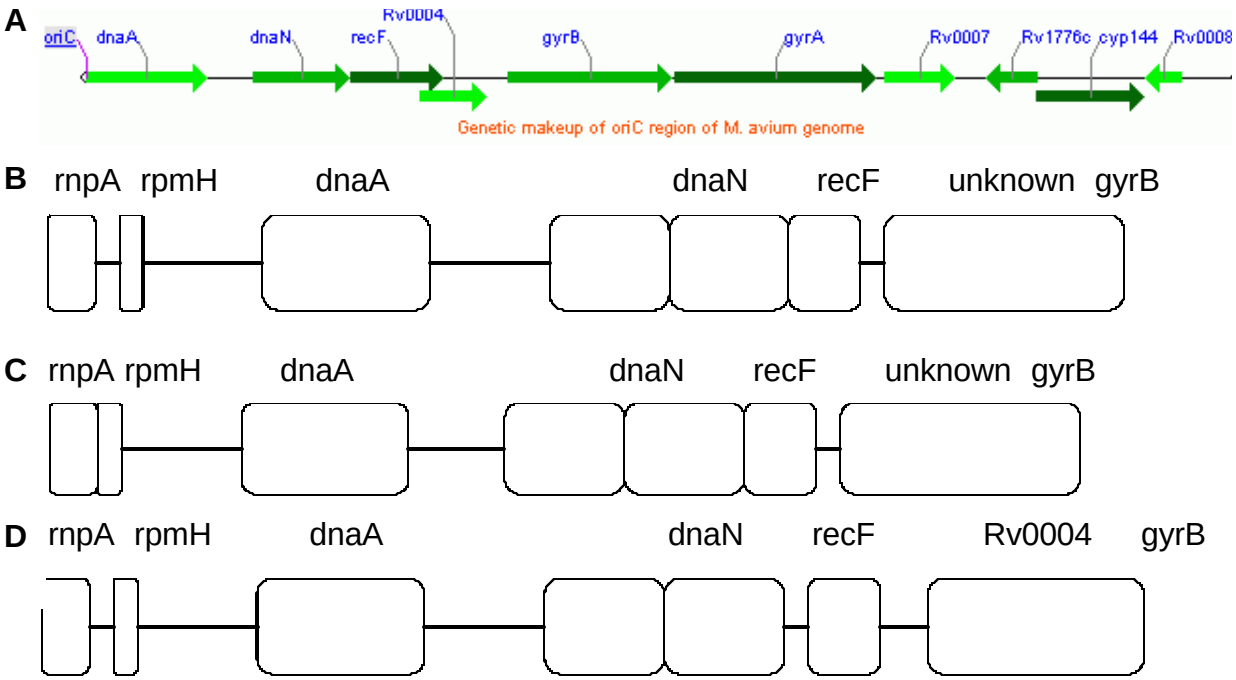


Figure 2. Diagram of the origin of replication (oriC) region of our annotation of the *M. avium subsp. avium* genome (a), compared with a published annotation of the same region for *M. avium subsp. avium* (b), *M. avium subsp. paratuberculosis* (c) and *M. tuberculosis* (d) (8) .

these automated steps within Microsoft Word™ or Microsoft Excel. As a final example, BlastParser™ (http://cbi.swmed.edu/computation/blastparser/) proved to be a tool of great value in the transfer of properly formatted text-based alignment result files. This software is able to convert BLAST output files into tab-delimited text that can be opened within Microsoft Excel™, thus simplifying BLAST result comparison between the databases. Table 2 lists some of the resources that we consider to be useful for future genomic projects.

Sequencing inaccuracies were encountered during the annotation process and highlighted an important limitation of bioinformatic work in general; the downstream analysis is only as good as the precision of raw data. The sequencing errors we were able to detect were noticed through careful comparison of the sequences of MAA and the M. tuberculosis genomes. For instance, in some cases consecutive ORFs overlapped in MAA while their M. tuberculosis counterparts were shorter and contiguous. By

Table 2. List of useful resources for annotation of bacterial genomes.

Resource	Description and Website
Artemis™	Genomic visualization software; The Sanger Institute: http://www.sanger.ac.uk/Software/Artemis/
Genamics Expression™	DNA and protein sequence analysis software; Genamics Inc.: http://genamics.com/expression/index.htm
National Center for Biotechnology Information (NCBI) Website	Online BLAST tools, including: http://www.ncbi.nlm.nih.gov/BLAST/ http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Genome
BlastParser™	A program that parses/reformats BLAST query results; Informatics Group at the Center for Biomedical Inventions: http://cbi.swmed.edu/computation/blastparser/
Microsoft Office™	Word Processor and Database software; Microsoft Corporation, www.microsoft.com
TubercuList Website	Genome database: *Mycobacterium tuberculosis* H37Rv; Pasteur Inst.: http://genolist.pasteur.fr/TubercuList
The Institute for Genomic Research(TIGR) website	The Institute for Genome Research, currently determining the Mycobacterium avium subsp. avium strain 104 sequence, http://www.tigr.org/
Mycobacterium avium comparative genomics	Comparative Genomics of the MAC at the Behr Lab, http://www.molepi.mcgill.ca/mac.htm

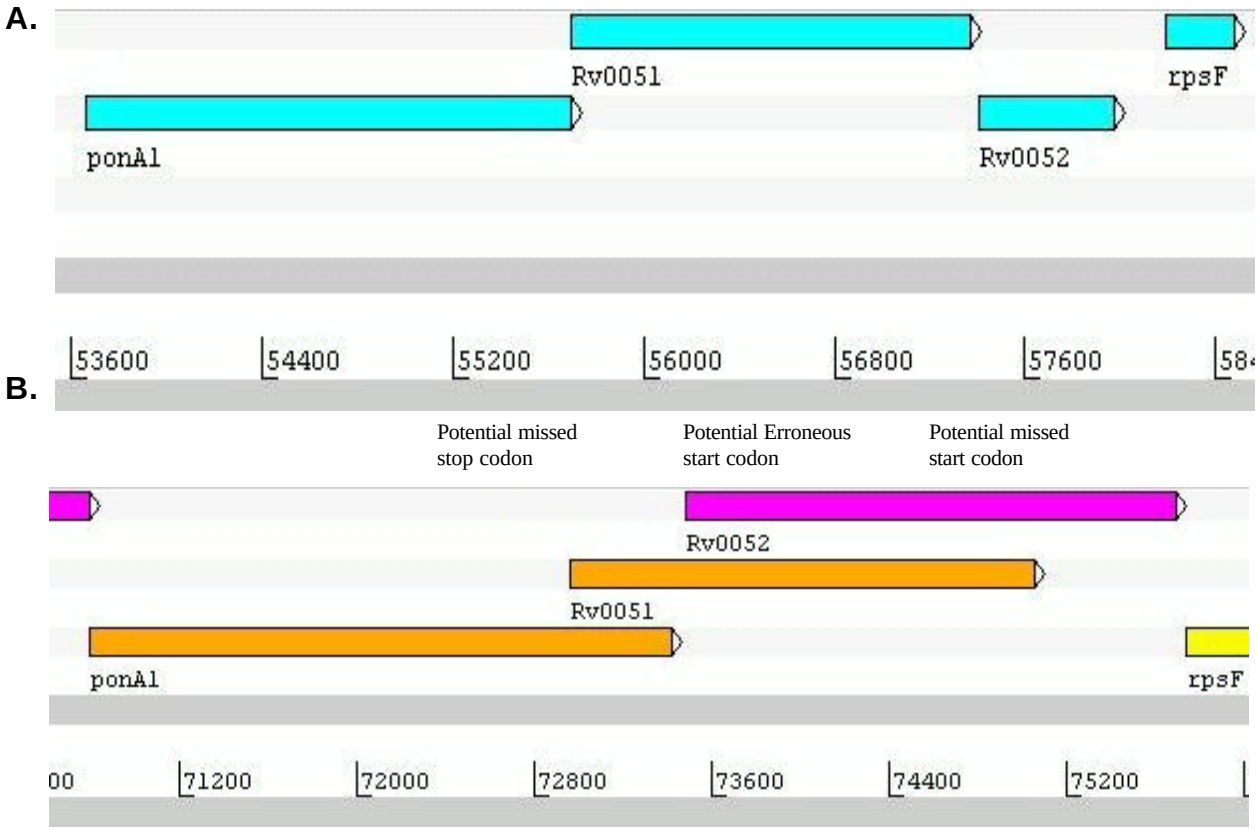


Figure 3. A possible sequencing error when comparing *M. tuberculosis* (a) and the complementary ORFs in *M. avium* subsp. avium (b). The ORFs in MAA overlap and are longer than the corresponding *M. tuberculosis* genes. These overlaps and gene length discrepancies are most likely due to sequencing inaccuracies, where stop and start codons are missed.

BLASTing only a part of these ORFs, we often found that high homology with the corresponding M. tuberculosis ORF was restricted to only part of the ORF, while the overlapping area would then show homology to an adjacent M. tuberculosis ORF. The observation that different ORFs were not distinguished in MAA suggested to us that sequencing errors at stop codons had obscured the predicted end of the ORF. Figure 3 shows an example of where a sequencing error occurred. The degree to which such errors impact on other aspects of the annotation unfortunately can only be determined when a "clean", final version of the MAA genome is released.

The annotation of the MAA 104 genome was intended for internal use, in order to facilitate genomic research of the MAC. In ongoing work, microarray-based genomic comparisons between members of the MAC have revealed large sequence polymorphisms, which are being used to construct a tentative phylogeny of the complex and to explore the biology of these organisms. Furthermore, the process and automations that were developed during the annotation project can be transported to annotating other sequenced genomes.

ACKNOWLEDGEMENTS

Funding for this project was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC). MK and MKL obtained research bursaries from the Fonds de la Recherche en Santé du Québec (FRSQ) and NSERC respectively. MB is a New Investigator of the Canadian Institutes of Health Research. MS is a recipient of the CIDS/CIHR/Bayer Healthcare fellowship award and is currently funded by the Fonds de la Recherche en Sante du Quebec (FRSQ). Preliminary sequence data was obtained from The Institute for Genomic Research website at http://www.tigr.org. We acknowledge Serge Mostowy for his original annotation of the MAA genome that served as an example for our efforts and provided the template for the DNA microarray design. We thank Melinda Pryor-Stinear of the Pasteur Institute for supplying us with Release 5 of the M. tuberculosis sequence and annotation well before its public release date, and Laura Carranza for insightful comments and discussions.

REFERENCES

1. Cole, S. T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. V. Gordon, K. Eiglmeier, S. Gas, C. E. Barry, III, F. Tekaia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, B. G. Barrell, and . 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393:537-544.
2. Behr, M. A., M. A. Wilson, W. P. Gill, H. Salamon, G. K. Schoolnik, S. Rane, and P. M. Small. 1999. Comparative genomics of BCG vaccines by whole-genome DNA microarray. Science 284:1520-1523.
3. Mostowy, S., D. Cousins, J. Brinkman, A. Aranaz, and M. A. Behr. 2002. Genomic deletions suggest a phylogeny for the *Mycobacterium tuberculosis* complex. J.Infect.Dis. 186:74-80.
4. Kato-Maeda, M., J. T. Rhee, T. R. Gingeras, H. Salamon, J. Drenkow, N. Smittipat, and P. M. Small. 2001. Comparing genomes within the species *Mycobacterium tuberculosis*. Genome Res. 11:547-554.
5. El Zaatari, F. A., M. S. Osato, and D. Y. Graham. 2001. Etiology of Crohn's disease: the role of *Mycobacterium avium* paratuberculosis. Trends Mol.Med. 7:247-252.
6. Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. J.Mol.Biol. 215:403-410.
7. Bannantine, J. P., E. Baechler, Q. Zhang, L. Li, and V. Kapur. 2002. Genome scale comparison of *Mycobacterium avium* subsp. paratuberculosis with *Mycobacterium avium subsp. avium* reveals potential diagnostic sequences. J.Clin.Microbiol. 40:1303-1310.
8. Bannantine, J. P., Q. Zhang, L. L. Li, and V. Kapur. 2003. Genomic homogeneity between *Mycobacterium* a*vium subsp. avium* and *Mycobacterium avium subsp. paratuberculosis* belies their divergent growth rates. BMC.Microbiol. 3:10.

David D. Shersher, Maksim Kirtsman, and **Mikael Katz-Lavigne** plan to undertake studies in medicine by the fall of 2004. They hope their undergraduate training in Microbiology/Immunology and their exposure to research will serve them well in a career combining clinical medicine and active research. **Dr. Makeda Semret** is a postdoctoral fellow and **Dr. Marcel A. Behr** is an Associate Professor at McGill University. They perform genomic studies of Mycobacteria with the aim to improve diagnosis and prevention of mycobacterial diseases.

ORIGINAL ARTICLE

Cost-Effectiveness of Positron Emission Tomography in Recurrent Colorectal Cancer in Canada

Jeffrey Scott Sloka, M.D., Ph.D.[†*], Peter Dorroch Hollett, M.D.[§], K. Maria Mathews, Ph.D.[¥]

BACKGROUND: Several studies over the past decade have demonstrated that 2-fluoro-2-D-[^{18}F]fluorodeoxyglucose (FDG) positron emission tomography (PET) is more accurate than computed tomography (CT) for the staging of recurrent colorectal carcinoma. This study uses quantitative decision tree modeling and sensitivity analysis to assess the cost-effectiveness of a PET-based management strategy for staging recurrent colorectal carcinoma in Canada. Both management costs and life expectancy are determined. METHODS: Two patient management strategies were compared - one using CT alone and one using both CT and PET. A survey of recent literature was used to construct a meta-analyses of available studies for the accuracy of PET in staging recurrent colorectal carcinoma. Life expectancies were determined from recent Canadian statistics, and expected life expectancies with disease were calculated from published survival rates. Management costs were determined from: estimates of the installation cost of PET facilities in Canada; management costs from our institutions; and recently published Canadian cost estimates of various procedures. RESULTS: A cost savings of $1,758 per person is expected for a PET and CT strategy, along with a slight increase in life expectancy (3.8 days), when compared with a CT alone strategy. This cost savings stemmed from avoided surgeries and remained in favour of the PET strategy when subjected to a rigorous sensitivity analysis.

INTRODUCTION

Colorectal cancer is a significant cause of morbidity and mortality in the Canadian population. In the year 2001, there were 17,000 new cases of colorectal cancer in Canada, and 6,500 people died from the disease (1). Colorectal cancer is also an expensive disease to diagnose and treat. Diagnostic tools such as CT imaging and colonoscopy are used to guide expensive treatment options such as chemotherapy and surgery. The accuracy of testing is essential for cost-effective decision making. Inaccuracies result in additional and inappropriate procedures (e.g. surgeries, colonoscopies) that both put the patient at an increased risk of procedure morbidity and mortality, and extra costs are incurred by the third party payer. The addition of a new, more accurate diagnostic test to the decision process benefits both the patient and the third party payer by reducing the number of inappropriate procedures.

Of the people who initially present with colorectal cancer and are resected for cure, 25-40% have disease recurrence, a proportion of whom are potentially curable by a second resection (Figure 1). Patients who present for detection and staging of recurrent colorectal cancer have a prevalence of recurrent disease of approximately 85% (2)(Table 1). The cost effective surveillance, diagnosis, and therapy for the recurrence of colorectal cancer depends on the accurate determination of who is appropriate for a second

* To whom correspondence should be addressed: Jeffrey Scott Sloka, MD, PhD, Faculty of Medicine, Memorial University of Newfoundland 108 Moss Heather Dr., St. John's, Newfoundland, A1B 4S1 e-mail: p97jss@mun.ca

† Faculty of Medicine, Memorial University of Newfoundland

§ Chief - Division of Nuclear Medicine, Department of Medical Imaging, Health Care Corporation of St. John's, St. John's, Newfoundland

¥ Assistant Professor Health Policy/Heath Care Delivery Division of Community Health, Memorial University of Newfoundland

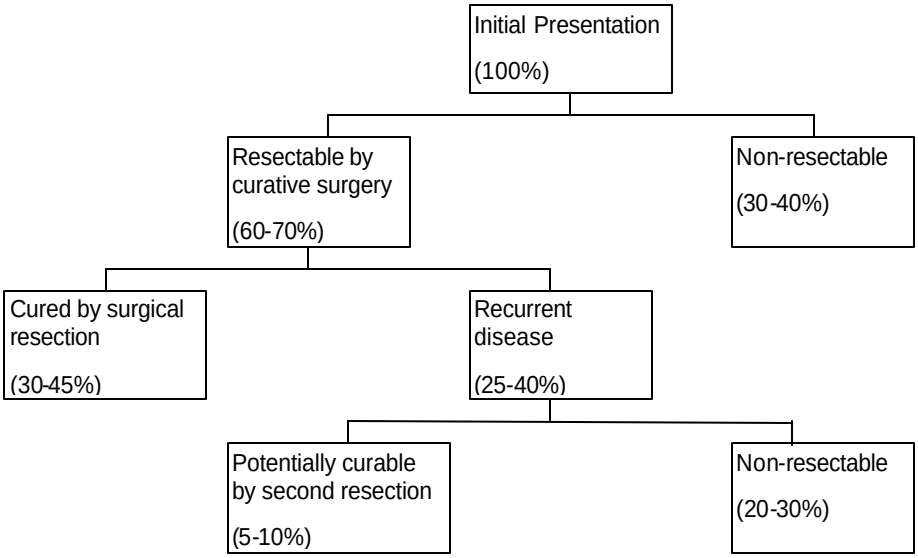


Figure 1. Epidemiological natural history of recurrent colorectal cancer. (Percents represent a proportion of those patients that initially present)(11). Sixty to seventy percent of new cases present with primary stage II or III disease, most of whom are potential candidates for curative surgical resection(11). After surgical resection, metastatic or locally invasive disease, primarily to the liver or lung, recurs in 35-40% of patients (25-40% of total presentations)(11;12), and approximately 25% of these recurrences are potentially curable. Only 29-46% of curative resection candidates are truly resectable at the time of surgery due to the underestimation of the extent of disease(12-20)

surgical resection. Preoperatively, a computed tomography (CT) study is ordered by the surgeon to aid in localization of recurrence.

Positron Emission Tomography (PET) utilizing [^{18}F]2-fluoro-2-deoxy-D-glucose (FDG) has been evaluated in several clinical studies for use in staging recurrent colorectal cancer and guiding treatment (2-8). PET using FDG is a noninvasive imaging modality that provides information useful for tumor imaging (9,10). Increased glucose utilization by malignant cells results in increased FDG uptake, which is used to localize metastatic sites. PET examines the entire body and can therefore identify sites of distant metastases in the pelvis, liver and lung. Studies have shown that PET may be more sensitive and specific than CT in detecting localized and metastatic recurrence of disease and therefore may be more accurate for directing surgical management (2-8).

The aim of this study is to compare two diagnostic strategies for recurrent colorectal cancer, one employing CT only and one employing PET and CT. The study will compare the marginal cost effectiveness of the addition of PET for the preoperative staging of recurrent colorectal cancer in Canada and will also compare the number of unnecessary surgeries performed between the two strategies. (Cost effectiveness analysis compares both economic costs/benefits and survival outcomes of proposed management strategies and expresses the results as a ratio of costs spent or saved to an outcome, e.g. number of life years saved (21)). The effects of the additional diagnostic tests are also compared in terms of the difference in life expectancies to determine any detriment to the patient. The viewpoint of the analysis is the hospital.

The cost effectiveness of using PET for the staging of recurrent colorectal cancer has been demonstrated by others in certain economic environments (22,23). However, the cost effectiveness of PET for recurrent colorectal cancer has not been studied for centers in Canada.

METHODS

Model structure

Decision models of established protocols from both the literature (5,24) and from current local practice were structured with two outcomes: cost and life expectancy (Figure 2), so that comparison of outcomes (total expected costs, expected life expectancy, and total number of surgeries) could be made. To construct the decision tree, the sensitivity and specificity values of each diagnostic test were identified. Mortality rates of each procedure were also determined as reported in the literature. The time horizon of the study was from the initial diagnostic studies to the final treatment modalities of the first disease recurrence.

The first decision tree uses CT to stage the patient's disease, while the second decision tree uses both PET and CT to stage the patient's disease. All results from the image modality (conservatively) have a colonoscopy performed with subsequent biopsy if positive. If negative, the disease is confirmed using biopsy or diagnostic laparotomy. All patients receive a

CT study in the PET strategy to provide the surgeon with anatomical information necessary for surgery. Every patient with a positive PET result, regardless of the CT result, is sent to biopsy for confirmation of disease (conservative treatment), thus ensuring that patients who are falsely positive of metastases receive the benefit of curative surgery. In the decision trees, a "+ve" in a particular test indicates that the patient has tested positive for metastatic disease. Further tests may be required to confirm this, at which time the patient may be a candidate for other therapies such as chemotherapy.

Population

In terms of life expectancy calculations, the representative population chosen for this study is a 65 year old person (an average age of presenting patients from some recent studies (25-29)), presenting with suspected recurrent colorectal cancer (having a mean survival of 16 months (30)), a proportion of whom are not candidates for resection due to disease extent. We used a theoretical sample of 1000 patients to estimate the number of procedures for each strategy.

Probabilities

Meta-analysis is the technique of combining the results of several studies to strengthen conclusions about individual studies when taken as a whole. Following guidelines outlined in the literature (31,32), a search was performed using: the MEDLINE keywords "PET, CT, and recurrent colorectal cancer"; published abstracts; and references noted in the above studies. Selected studies included both retrospective and prospective studies (2-8). Studies that did not publish the numbers used to derive the sensitivity and specificity were not included in the meta-analysis because combined sensitivities and specificities could not then be calculated. Only studies that confirmed the diagnosis with biopsy were included, and because this study determines outcomes based on the patient, any study that published sensitivities and specificities based strictly on the number of identified lesions was not included (33). No tests of homogeneity were used, however Figure 3 indicates the study results visually.

The aggregate average of eight studies was used to calculate the model sensitivities and specificities (2-7, 34). These studies did not have an equal average age of patients, and some did not blind the results between the PET and CT studies. Since a stricter set of inclusion criteria results in smaller numbers of patients in the meta analysis, most studies were included in determining the aggregate average with the knowledge that some variability across diagnostic usage and patient population would exist. It was noted, however, that all included studies showed PET to be more sensitive and specific within each study.

Table 1. Parameters used in the decision model. The derivation of all parameters appears throughout the methods section. The disease prevalence is the prevalence of recurrent colorectal cancer in the population of patients presenting for diagnosis of recurrence.

Population		**Biopsy**	
Disease prevalence	**85%**	Cost	$118
		Mortality	0.7%
CT		Sensitivity	100%
Cost	$462	Specificity	100%
Mortality	0.0025%		
Sensitivity	76.2%	**Surgical Resection**	
Specificity	69.4%	Cost	$16 479
		Mortality	6.5%
PET			
Cost	$1029	**Life expectancy**	
Mortality	0%	Local population	17.4 yrs
Sensitivity	93.3%	Patients with stage IV	1.1 yrs
Specificity	92.7%	Patients after curative surgery	2.6 yrs
Colonoscopy			
Cost	$168	**Chemotherapy**	
Mortality	0.005%	Cost	see text
Sensitivity	93%		
Specificity	95%		

Alteration of surgical management

The use of FDG-PET has been reported to alter the management of patients prior to surgery given the additional accuracy PET provides over CT staging (2,4-8, 35-37). Alteration of management may include: the detection of unknown liver metastases which may augment surgical management; the detection of extensive metastases which may exclude the possibility of surgery and direct the management towards chemotherapy and/or palliative care; or the addition of surgical management if recurrent cancer is confirmed

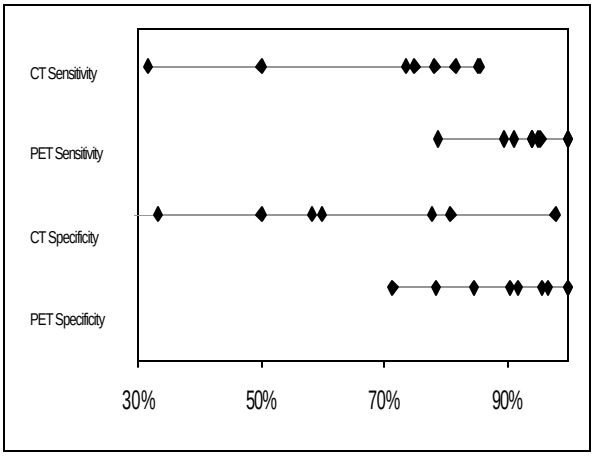


Figure 3. CT and PET sensitivities and specificities for 8 studies used in the meta analysis plotted to demonstrate clustering

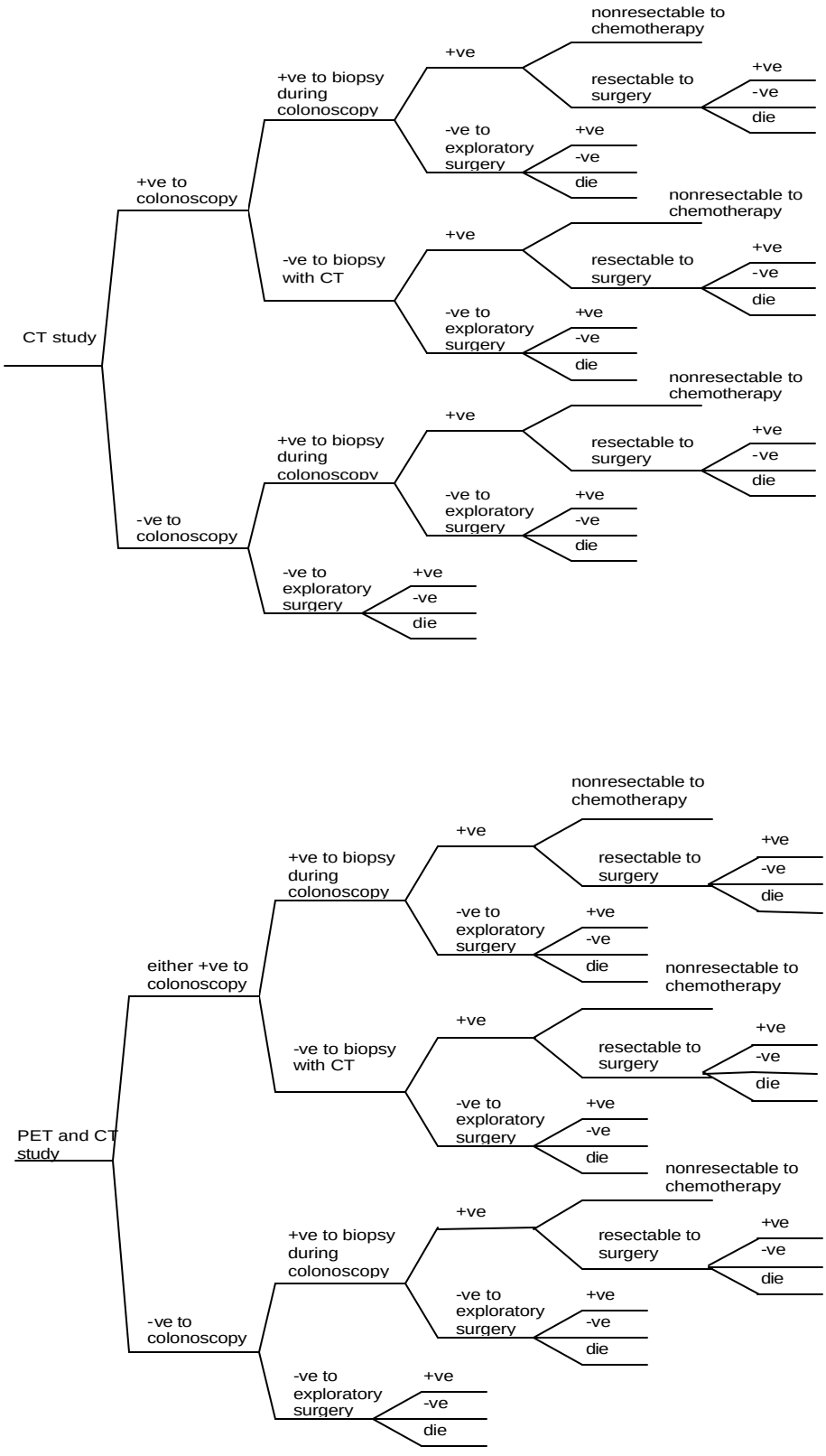


Figure 2 - Decision tree for both PET and CT strategies. The outcomes of each test follow the test intersection, and each outcome is assigned a probability of occurring given the sensitivity and specificity of the test for that given block(33). The "+ve" in a particular test in the decision trees indicates that the patient has tested positive for disease. (5)

where other tests failed to detect its existence. The combined alterations in surgical management for 10 studies where the detection of extensive metastases excluded the possibility of surgery resulted in a combined 28.7% of surgical candidates that were not eligible for surgery because of extensive metastases. These patients were subjected to unnecessary surgery (with an associated mortality, morbidity, and cost), whereas different treatment options would have been offered if better staging was available.

Mortality data

The associated risk of the various procedures are included in the decision tree model because they affect the life expectancy outcome of the patient. The mortality associated with CT is primarily attributable to the intravenous administration of contrast material, reported to be 0.0025% (1 in 40 000) (38), and was chosen as a baseline mortality for this study (39,40). No adverse reactions or complications due to the administration of FDG have been reported to date (39,40).

The risk of mortality associated with the surgical resection of colon or rectum has been reported in several studies (41-48). There is wide variation in the report of several smaller studies; one large study from Australia (48) reported a perioperative mortality rate of 6.5%, which was used for this study.

The reported complication rate of colonoscopy with biopsy is 0.2% (mainly due to perforation of the bowel), resulting in an approximate 0.005% mortality rate (49). Liver biopsy is also performed to confirm the presence of metastases, and the mortality from various large combined series is approximately 0.01% (50). Approximately 84% of recurrences include the colon or rectum, and approximately 12% of all recurrences metastasize only to the liver (11,12). In our practice, all suspected colorectal recurrences would be biopsied, as well as all isolated suspected liver metastases. Therefore, this study used a weighted average mortality for biopsy of 0.84(0.005%) + 0.12(0.01%), or 0.0054%.

Life expectancy

The life expectancy of an average 65 year old Canadian is 18.3 years. The 5 year survival rate of people with recurrent colorectal cancer was recently reported to be 19.2% (a weighted average based on stage of disease) in a meta-analysis of several follow-up studies (51). Using the DEALE method for determining the 5 year survival rates of disease (52, 53), the life expectancy calculated for all patients with unstaged recurrent colorectal cancer is 2.6 years. The mean survival of someone with untreated extensive colorectal metastasis is 13.1 months, whereas the mean survival of those treated with chemotherapy is 16.3 months (30).

Costs

Our current fee schedules and cost accounting systems along with the medical literature were surveyed to obtain the most recent Canadian values for all procedure costs and outcomes (Table 1). Year 2000 Canadian dollars was used as the currency, correcting for inflation using the Canadian Consumer Price Index (54). Costs outside the time horizon of the study, costs to society such as lost productivity, and indirect costs to quality of life were not used.

In this study, the cost of capital equipment is discounted over the expected lifetime of the equipment using a standard annuity formula, amortized over the equipment lifetime at an assumed interest rate of 6%. Estimated equipment costs are outlined in Table 2. The estimated equipment lifetime for a positron tomograph is 5 years, and it is 10 years for a cyclotron installation (23).

For both a PET camera and a cyclotron, the estimated yearly operating cost is $1 625K. Assuming that each PET installation operated at full potential on a 1 shift per day basis (7 patients per shift), there would be a yearly capacity for 1 750 cases. The total cost per case would be $1625K/1750cases = $929/case. Allowing for a physician remuneration of $100 per case (an estimate based on other modalities), the total cost per case is estimated to be $1029. Capital acquisition, depreciation, and annual operating estimates were not included for CT, because one CT study is performed per person in each strategy, and these costs cancel out when the two management strategies are compared. Overhead costs were not included in this analysis.

The cost of surgery for the resection of recurrent disease is calculated as an aggregate of bed costs and professional charges (Table 3). Those patients who undergo hepatic resection usually stay in hospital for 2 weeks, 5 days of which are spent in ICU. Those who undergo colorectal resection usually spend 1 to 2 weeks in a ward bed. Up to 35% of patients have resectable liver metastases. An average total cost for resection was estimated to be $16,479.34. The cost of a thorax, abdomen and pelvic CT exam was estimated to be $200 (55) plus $262 professional costs. Professional costs for CT-guided biopsy are $118, and $168 for colonoscopy.

Estimates of the cost of chemotherapy for colorectal cancer were not available, although estimates of chemotherapy for lung cancer are known (56) and were used to roughly estimate the costs of chemotherapy to be $10,000.

Analysis

Expected costs and outcomes were calculated for each of the decision models using a standard decision analysis software package (57). The probability of

Table 2. PET Camera Capital and Operating Cost Estimates

Capital Costs		Operating Costs	
Conventional Positron Tomograph	**$2 500 000**	**Salaries for 2 camera technologists**	**$93 000**
Cyclotron	**$2 400 000**	**Part time for a supporting scientist**	**$28 000**
		Camera supplies	**$100 000**
		Cyclotron salaries for chemist, operator and radiopharmacist	**$175 000**
		Cyclotron supplies	**$20 000**
		Cyclotron utilities	**$15 000**
		18F targets	**$50 000**
		Cyclotron maintenance	**$225 000**

Table 3. Cost structure for surgical resection..
* represents hospital costs that were assumed to have been combined into bed cost per day.

Surgery	
Preoperative testing	
EKG	$63.25
Chest X ray	$10.55
Blood tests	*
Preoperative consult	
Surgeon	$49.92
Preadministrative clinic	
Nurse	*
Medical problems -> specialist	Variable
Anesthesiologist consult	$33.50
Operating room	
Equipment	*
Facilities	*
Anesthesiologist fees	$149.52
Surgeon fees	$534.65
	$232.87
Nurse assistant fees	*
Surgical assistant fees	$32.04
Additional person	$32.04
Intraoperative ultrasound to assess liver	$17.69
Recovery room nurse	*
Recovery room costs	*
Total	$1,156.03

ICU		**Ward**	
Bed cost per day	$1,345.00	Bed cost	
ICU Specialist costs		per day	$840.00
First day	$100.39		

outcome at each decision step was derived from assigned prevalences, sensitivities, and specificities (33). Total expected costs and outcomes were calculated given the probability of outcome at each node (33). Sensitivity analysis was performed on all key variables to analyze sensitivity of results to inaccuracies of parameter estimates.

RESULTS

Given our theoretical sample of 1000 patients presenting for diagnostic evaluation and treatment, 580 surgeries would be performed with the CT model and 455 surgeries would be performed with the PET model. 125 people would avoid unnecessary surgery with the PET model compared to the CT model, due to a reduction in false positives from inaccurate staging. These patients would be sent directly to alternate forms of therapy and/or palliation without the additional risks of surgery. This large difference is due to the more accurate staging with the PET modality, but it significantly depends on the accuracy of biopsy in this model, conservatively assumed to be 100%. This difference also depends on the mutual exclusivity of the two diagnostic imaging modalities, which will never be the case in practice.

The expected cost of the CT alone strategy was $9,523 per person, and the expected cost of the PET strategy was $7,765 per person, translating to an expected savings of $1,758 per person using the PET strategy. This savings is associated with an increase in life expectancy (3.8 days). These results are due to the improved staging of recurrent colorectal cancer prior to surgery; patients with inoperable metastatic disease are directed away from surgery, a procedure associated with a high cost and mortality rate. Since the change in life expectancy is 3.8 days (in favour of the PET model), cost effectiveness calculations were not performed due to clinical insignificance. However, this change in life expectancy also demonstrates that the addition of PET into the decision model is not detrimental to the health of the patient.

Sensitivity analysis was performed on key parameters (such as the cost of an FDG-PET study or the specificity of a CT study) to determine the sensitivity of the cost savings and life expectancy to variation of these inputs (Table 4). For example, the expected cost per person is lower using the PET strategy unless the cost of a PET study increases beyond $2,787. Table 4 shows the limits at which the PET strategy has an expected cost per person less than the CT strategy, and Table 10 shows the limits at which the PET strategy demonstrates a better life expectancy per person than the CT strategy.

The sensitivity analysis shows that if the sensitivity and specificity of PET is reduced to 74.8% and 67.3% respectively, the CT only strategy has a lower expected cost. This is because the accuracy of PET has been reduced to the accuracy of CT, making the accuracy of staging for both models similar. The percentage of people that avoided surgery was also subjected to sensitivity analysis and it was determined that the PET strategy has a lower expected cost if the percentage of

Table 4. Sensitivity analysis for variables of interest used in the decision model. The variables of interest were varied until the expected cost/life expectancy of the CT strategy became less than the expected cost/life expectancy of the PET strategy. The "any" signifies that for any value of the given variable, the PET strategy is more cost effective than the CT strategy.

Variable of Interest	Cost Effectiveness	Life Expectancy	Baseline Value
Disease prevalence	>22.4%	>17.5%	85%
CT Cost	any		$462
PET Cost	<$2787		$1029
Surgery Cost	>$2922		$16 479
Biopsy Cost	any		$118
Chemotherapy Cost	<$100 000		$10 000
CT sensitivity	<87.3%	<91.8%	76.2%
CT specificity	any	any	69.4%
PET sensitivity	>73.8%	>44.3%	92.9%
PET specificity	>65.3%	>51.2%	93.4%
Avoidance of surgery	>3.2%	>11.3%	28.7%
Non-resectable	<95.0%	<71.6%	54-71% (13-16)

people that avoid surgery is greater than 3.2% of the people that are considered for curative surgery, which is much less than the value that was determined through the meta-analysis (28.7%). As well, for the PET strategy to be more cost effective than the CT strategy, less than 95.0% of patients should have non-resectable disease.

If the cost of surgery was reduced to less than $2922, the CT strategy would have a lower expected cost. Since the estimated cost of surgery is $16,479, the sensitivity for variations in this cost is low. However, if the screening procedures were not sufficiently accurate and they reduced the prevalence of the disease in the population of people presenting for diagnostic test at this level to less than 22.4% (our study estimate is 85%), the CT strategy would become more cost effective in terms of expected cost per patient. The cost of chemotherapy was varied between 0$ and $100K and the PET strategy was economical for this entire range.

DISCUSSION

The cost of medical care in Canada continues to rise, and now stands at $100 billion per year, approximately 9.3% of the gross domestic product(58). Due to the high cost of health care and the expectations of the Canadian people for a quality health care system with the best available diagnostic tools, clinical decision-making should also consider the cost of these decisions. It is desirable to find evidence-based clinical decision-making strategies where both a clinical and economic benefit coexist. Using a theoretical sample of 1000 patients and a decision tree analysis, we found that CT+PET were a cost-effective approach to determining the management of recurrent colorectal cancer. The additional PET study cost is more than compensated for by the savings realized from avoided surgeries. Over a range of values for the procedural parameters, the PET+CT strategy is shown to be more cost effective than the CT only strategy. There is a cost benefit of approximately $1,758 per person without a reduction in life expectancy.

Patients may benefit in several ways from the additional accuracy of staging techniques. Patients are fearful of recurrence (having been through the treatment of cancer at least once already) and accurate detection and staging contribute to their peace of mind. As well, morbidity and mortality are reduced through the avoidance of inappropriate surgery.

This study did not include the use of PET to stage primary colorectal cancer on initial presentation, although successful clinical studies have been performed to determine the usefulness PET in staging primary colorectal cancers (59) and for the evaluation of metastases to the liver (4,6).

A key assumption of this model is that this is a theoretical sample of 1000 patients who are at an average age for recurrent colorectal cancer. The accuracy of CT and PET may depend on the age of the population under study. This model was also constructed using our local practices which may or may not be similar to other practices elsewhere.

Another key assumption was that a "critical mass" of patients is needed for full utilization of a PET camera and cyclotron. Our calculations were based on full utilization; however, some centers in Canada do not have sufficient patient demand in terms of recurrent colorectal cancer to run a single shift daily for five days a week. Regions with underutilization would incur a higher cost than those without due to equivalent capital and yearly operating costs, but lower savings from fewer surgeries avoided. More study is required to determine regional options for centers with smaller catchments.

Selection of clinical studies for the meta analysis did not exclude studies that did not blind their results between CT and PET(39,40). In these cases, many patients are selected for PET imaging because of positive CT findings, thus introducing a case-selection bias resulting in an over-representation of positive CT findings (both false and true), with a concomitant overestimation of CT sensitivity and underestimation of specificity. Therefore, the CT and PET tests are not conditionally independent and some results do depend on diagnostic sequence. This does not, however, invalidate the meta analysis because the purpose is to determine the impact of adding a new modality to the patient's diagnostic algorithm, and the final test does

demonstrate accuracy (2). The clinical question for several studies centered on the detection of liver metastases, and for other studies the clinical question measured the detection of the recurrence of localized disease. Since both types of information are used to stage the disease, and since the staging of disease directs the management of the disease, both of these types of studies were included. It might be argued that the combination will increase the total accuracy because results from one modality may direct another.

Several assumptions were made to facilitate procedural modeling. The assumption that the biopsy is 100% accurate is not valid - biopsy may miss some lesions due to sampling error. Total costs, including clinician costs, blood tests, and other screening costs, were not included in this analysis but, similar to the minimal effect of CT cost variability on the expected cost savings, additional matched costs to both strategies may cancel out. At present, the use of PET technology has some drawbacks including the reliability of diagnosing tumors that are less than 1 cm^3. The meta-analysis is not as strict as it could be. The inclusion of several studies with clinical designs that are not 100% compatible is not optimal, but it does provide a more generalized approach for determining the overall accuracy of these diagnostic methods. At the present time, only one prospective study (2) has been published.

Overhead costs, palliative care costs and costs associated with quality of life were not included in this model. The overhead costs associated with operating a new PET center are unknown. It is assumed that these costs would be taken over by the imaging department and that shared costing within the department would reduce the costs incurred by each camera. The cost of chemotherapy was also not known and had to be estimated.

People who require palliation and who are sent to inappropriate resection surgery due to incorrect staging eventually require palliative care. As well, people that are falsely negative for recurrence yet have non-resectable disease may eventually present for palliation. Therefore, we assumed that those people who may require palliation will eventually receive palliation and so these costs do not change with the addition of a new diagnostic modality. This assumption may not be accurate because false negatives may present later for palliation (at a reduced total cost of palliation). False negative are reduced with the addition of PET. Therefore, a higher palliation cost may be realized for the addition of a more accurate diagnostic tool.

Quality of life can be represented in terms of actual costs (QALY). These costs were not included in this study. However, reducing the number of surgeries in people that do not require surgery would theoretically increase the quality of life in people that are determined to make the most out of life while they can.

The theoretical population sample may not reflect the actual population that presents for diagnosis. A sample of an average age of people presenting for diagnosis (a 65 year old) may not be representative of a regional population. As well, the accuracy of diagnosis of PET and CT could change with different population age (since the index of suspicion for recurrence may change). However, the change in life expectancy would be similar for any age since the change is strictly due to the reduction in the number of deaths due to surgery. A study based on real patient data would provide a more accurate representation.

CONCLUSIONS

Colorectal cancer is an important cause of morbidity and mortality in Canada; it is the 3rd most common cancer and the 3rd most common cause of cancer death. 17 000 new cases of colorectal cancer are expected this year and the lifetime risk of colorectal cancer is 6%. This study has described the use of PET technology for the staging of recurrent colorectal carcinoma, and may benefit patients in terms of a minor increase in life expectancy of 3.8 days, but it could also benefit the Canadian people in terms of reduced health care costs. This was shown to be true even for a wide variation in approximated variables used by this analysis. This study indicates that PET may be used economically in Canada in certain clinical situations.

REFERENCE

1. National Cancer Institute of Canada. Canadian Cancer Statistics 2000, Toronto, Canada, 2000. 2001.
2. Valk PE, Abella-Columna E, Haseman MK, Pounds TR, Tesar RD, Myers RW et al. Whole-body PET imaging with [^{18}F]fluorodeoxyglucose in management of recurrent colorectal cancer. Arch Surg 1999; 134(5):503-511.
3. Schiepers C, Penninckx F, De Vadder N, Merckx E, Mortelmans L, Bormans G et al. Contribution of PET in the diagnosis of recurrent colorectal cancer: comparison with conventional imaging. Eur J Surg Oncol 1995; 21(5):517-522.
4. Vitola JV, Delbeke D, Sandler MP, Campbell MG, Powers TA, Wright JK et al. Positron emission tomography to stage suspected metastatic colorectal carcinoma to the liver. Am J Surg 1996; 171(1):21-26.
5. Delbeke D, Vitola JV, Sandler MP, Arildsen RC, Powers TA, Wright JK, Jr. et al. Staging recurrent metastatic colorectal carcinoma with PET. J Nucl Med 1997; 38(8):1196-1201.
6. Boykin KN, Zibari GB, Lilien DL, McMillan RW, Aultman DF, McDonald JC. The use of FDG-positron emission tomography for the evaluation of colorectal metastases of the liver. Am Surg 1999; 65(12):1183-1185.
7. Fong Y, Saldinger PF, Akhurst T, Macapinlac H, Yeung H, Finn RD et al. Utility of 18F-FDG positron emission tomography scanning on selection of patients for resection of hepatic colorectal metastases. Am J Surg 1999; 178(4):282-287.
8. Flanagan FL, Dehdashti F, Ogunbiyi OA, Kodner IJ, Siegel BA. Utility of FDG-PET for investigating unexplained plasma CEA

elevation in patients with colorectal cancer. Ann Surg 1998; 227(3):319-323.
9. Hoh CK, Schiepers C, Seltzer MA, Gambhir SS, Silverman DH, Czernin J et al. PET in oncology: will it replace the other modalities? Semin Nucl Med 1997; 27(2):94-106.
10. Conti PS, Lilien DL, Hawley K, Keppler J, Grafton ST, Bading JR. PET and [^{18}F]-FDG in oncology: a clinical update. Nucl Med Biol 1996; 23(6):717-735.
11. Desch CE, Benson AB, III, Smith TJ, Flynn PJ, Krause C, Loprinzi CL et al. Recommended colorectal cancer surveillance guidelines by the American Society of Clinical Oncology. J Clin Oncol 1999; 17(4):1312.
12. August DA, Ottow RT, Sugarbaker PH. Clinical perspective of human colorectal cancer metastasis. Cancer Metastasis Rev 1984; 3(4):303-324.
13. Grage TB, Vassilopoulos PP, Shingleton WW, Jubert AV, Elias EG, Aust JB et al. Results of a prospective randomized study of hepatic artery infusion with 5-fluorouracil versus intravenous 5-fluorouracil in patients with hepatic metastases from colorectal cancer: A Central Oncology Group study. Surgery 1979; 86(4):550-555.
14. Macdonald JS, Kisner DF, Smythe T, Woolley PV, Smith L, Jr., Schein PS. 5-Fluorouracil (5-FU), methyl-CCNU, and vincristine in the treatment of advanced colorectal cancer: phase II study utilizing weekly 5-FU. Cancer Treat Rep 1976; 60(11):1597-1600.
15. Buroker T, Kim PN, Groppe C, McCracken J, O'Bryan R, Panettiere F et al. 5FU infusion with mitomycin-C versus 5 FU infusion with methyl-CCNU in the treatment of advanced colon cancer: a Southwest Oncology Group Study. Cancer 1978; 42(3):1228-1233.
16. Kemeny N, Younes A, Seiter K, Kelsen D, Sammarco P, Adams L et al. Interferon alpha-2a and 5-fluorouracil for advanced colorectal carcinoma. Assessment of activity and toxicity. Cancer 1990; 66(12):2470-2475.
17. Cohen AM, Minsky BD, Schilsky RL.In: De Vita VT,Jr. Hellmann S, Rosenberg SA (eds). Cancer: principles and practice of oncology, 4th ed. Lippincott, Philadelphia, p929.
18. Taylor I. Liver metastases from colorectal cancer: lessons from past and present clinical studies. Br J Surg 1996; 83(4):456-460.
19. Foster JH, Lundy J. Liver Metastases. Curr Probl Surg 1981; 18(3):157-202.
20. Fong Y, Salo J. Surgical therapy of hepatic colorectal metastasis. Semin Oncol 1999; 26(5):514-523.
21. Johannesson M. On the estimation of cost-effectiveness ratios. Health Policy 1995; 31(3):225-229.
22. Dsetlein M, Sheidhauer K, Lautenbach K, Schicha H. Quality criteria for cost-benefit analysis in oncological nuclear medicine. Zeitschrift fur Arzlich Fortbildung und Qualitatssichanung 93[1], 49-55. 2001.
23. Beanlands R, Chamberlain M, Cripps C, et al. Position paper on the future of positron emission tomography in Ontario. Council of Medical Imaging, Oakville, Ontario . 2001. 1999.
24. Institute for Clinical PET. Clinical application and economic implications of PET in the assessment of colorectal cancer recurrence: A retrospective study. ICP International Conference 2001.
25. Chiappa A, Zbar AP, Biella F, Staudacher C. Survival after repeat hepatic resection for recurrent colorectal metastases. Hepatogastroenterology 1999; 46(26):1065-1070.
26. Law WL, Chu KW, Ho JW, Tung HM, Law SY, Chu KM. Self-expanding metallic stent in the treatment of colonic obstruction caused by advanced malignancies. Dis Colon Rectum 2000; 43(11):1522-1527.
27. Cheung PC, Mackillop WJ, Dixon P, Brundage MD, Youssef YM, Zhou S. Involved-field radiotherapy alone for early-stage non-small-cell lung cancer. Int J Radiat Oncol Biol Phys 2000; 48(3):703-710.
28. Hoffman PC, Mauer AM, Vokes EE. Lung cancer. Lancet 2000; 355(9202):479-485.
29. Knott-Craig CJ, Howell CE, Parsons BD, Paulsen SM, Brown BR, Elkins RC. Improved results in the management of surgical candidates with lung cancer. Ann Thorac Surg 1997; 63(5):1405-1409.
30. Durand-Zaleski I, Roche B, Buyse M, Carlson R, O'Connell MJ, Rougier P et al. Economic implications of hepatic arterial infusion chemotherapy in treatment of nonresectable colorectal liver metastases. Meta-Analysis Group in Cancer. J Natl Cancer Inst 1997; 89(11):790-795.
31. Oxman AD, Cook DJ, Guyatt GH. Users' guides to the medical literature. VI. How to use an overview. Evidence-Based Medicine Working Group. JAMA 1994; 272(17):1367-1371.
32. Gambhir SS. Decision analysis in nuclear medicine. J Nucl Med 1999; 40(9):1570-1581.
33. Sox HC. Medical decision making. Butterworths . 2001. Boston.
34. Akhurst T, Larson SM. Positron emission tomography imaging of colorectal cancer. Semin Oncol 1999; 26(5):577-583.
35. Beets G, Penninckx F, Schiepers C, Filez L, Mortelmans L, Kerremans R et al. Clinical value of whole-body positron emission tomography with [^{18}F]fluorodeoxyglucose in recurrent colorectal cancer. Br J Surg 1994; 81(11):1666-1670.
36. Lai DT, Fulham M, Stephen MS, Chu KM, Solomon M, Thompson JF et al. The role of whole-body positron emission tomography with [^{18}F]fluorodeoxyglucose in identifying operable colorectal cancer metastases to the liver. Arch Surg 1996; 131(7):703-707.
37. Imdahl A, Reinhardt MJ, Nitzsche EU, Mix M, Dingeldey A, Einert A et al. Impact of ^{18}F-FDG-positron emission tomography for decision making in colorectal cancer recurrences. Langenbecks Arch Surg 2000; 385(2):129-134.
38. Hartman GW, Hattery RR, Witten DM, Williamson B, Jr. Mortality during excretory urography: Mayo Clinic experience. AJR Am J Roentgenol 1982; 139(5):919-922.
39. Gambhir SS, Hoh CK, Phelps ME, Madar I, Maddahi J. Decision tree sensitivity analysis for cost-effectiveness of FDG-PET in the staging and management of non-small-cell lung carcinoma. J Nucl Med 1996; 37(9):1428-1436.
40. Kosuda S, Ichihara K, Watanabe M, Kobayashi H, Kusano S. Decision-tree sensitivity analysis for cost-effectiveness of chest 2- fluoro-2-D-[(^{18})F]fluorodeoxyglucose positron emission tomography in patients with pulmonary nodules (non-small cell lung carcinoma) in Japan. Chest 2000; 117(2):346-353.
41. Bohm B, Grundel K, Neudecker J, Schwenk W. [Can colorectal carcinomas be resected laparoscopically? Technically possible, long-term outcome is still uncertain]. MMW Fortschr Med 2000; 142(8):30-32.
42. Baccari P, Di Palo S, Redaelli A, Carlucci M, Staudacher C. [Laparoscopic versus conventional surgery in the treatment of colorectal diseases]. Chir Ital 2000; 52(1):17-27.
43. Bolton JS, Fuhrman GM. Survival after resection of multiple bilobar hepatic metastases from colorectal carcinoma. Ann Surg 2000; 231(5):743-751.
44. Imamura H, Kawasaki S, Miyagawa S, Ikegami T, Kitamura H, Shimada R. Aggressive surgical approach to recurrent tumors after hepatectomy for metastatic spread of colorectal cancer to the liver. Surgery 2000; 127(5):528-535.
45. Schiedeck TH, Schwandner O, Baca I, Baehrlehner E, Konradt J, Kockerling F et al. Laparoscopic surgery for the cure of colorectal cancer: results of a German five-center study. Dis Colon Rectum 2000; 43(1):1-8.
46. Rosen SA, Buell JF, Yoshida A, Kazsuba S, Hurst R, Michelassi F et al. Initial presentation with stage IV colorectal

cancer: how aggressive should we be? Arch Surg 2000; 135(5):530-534.
47. Fernandez Merino FJ, Vicente LE, Nuno GJ, Zarzosa G, Barrasa SA, Meneu JC. Resection of colorectal liver metastases: analysis of risk factors. Rev Esp Enferm Dig 2000; 92(2):97-104.
48. Ansari MZ, Collopy BT, Hart WG, Carson NJ, Chandraraj EJ. In-hospital mortality and associated complications after bowel surgery in Victorian public hospitals. Aust N Z J Surg 2000; 70(1):6-10.
49. Hart R, Classen M. Complications of diagnostic gastrointestinal endoscopy. Endoscopy 1990; 22(5):229-233.
50. Sherlock S DJ. Diseases of the liver and biliary system, 10th ed. London: Blackwell Science . 2001.
51. Kievit J. Colorectal cancer follow-up: a reassessment of empirical evidence on effectiveness. Eur J Surg Oncol 2000; 26(4):322-328.
52. Beck JR, Kassirer JP, Pauker SG. A convenient approximation of life expectancy (the "DEALE"). I. Validation of the method. Am J Med 1982; 73(6):883-888.
53. Beck JR, Pauker SG, Gottlieb JE, Klein K, Kassirer JP. A convenient approximation of life expectancy (the "DEALE"). II. Use in medical decision-making. Am J Med 1982; 73(6):889-897.
54. Evans WK, Will BP, Berthelot JM, Wolfson MC. The economics of lung cancer management in Canada. Lung Cancer 1996; 14(1):19-29.
55. Investigation for mediastinal disease in patients with apparently operable lung cancer. Canadian Lung Oncology Group. Ann Thorac Surg 1995; 60(5):1382-1389.
56. Coy P, Schaafsma J, Schofield JA. The cost-effectiveness and cost-utility of high-dose palliative radiotherapy for advanced non-small-cell lung cancer. Int J Radiat Oncol Biol Phys 2000; 48(4):1025-1033.
57. Gambhir SS, Gupta P, Shepherd JE, Allen MA, Hoh C, Maddahi J et al. MD@: a physician-friendly decision analysis tool. MD Comput 1998; 15(1):40-48.
58. Canadian Institute for Health Information. Health Care in Canada. ISBN: 1-55392-018-X. 9-5-2002.
59. Abdel-Nabi H, Doerr RJ, Lamonica DM, Cronin VR, Galantowicz PJ, Carbone GM et al. Staging of primary colorectal carcinomas with fluorine-18 fluorodeoxyglucose whole-body PET: correlation with histopathologic and CT findings. Radiology 1998; 206(3):755-760.

Scott Sloka is the first year neurology resident at MUN. He holds a BASc in Electrical Engineering and a PhD in Image Processing from theUniversity of Waterloo and an MD from Memorial University of Newfoundland. His research interests include image understanding, pattern recognition, autoimmunity and neuroepidemiology.
Peter Hollett is the chief of Nuclear Medicine at the Health Care Corporation for St. John's. His research interests include nuclear cardiology, respiratory disease and endocrine disorders.
Maria Mathews is an Assistant Professor of Health Policy/Health Care Delivery in the Division of Community Health, Faculty of Medicine, Memorial University of Newfoundland. She has a PhD in Health Policy, Management, and Evaluation from the University of Toronto and a Masters in Health Services. Administration from the University of Alberta. Her research interests include access issues in cancer care and primary care, and improving evidence based practice in health services management.

REVIEW ARTICLE

An Evaluation of Human Immunodeficiency Virus Pathogenicity and Treatment Using Glycobiology

Aleksandra Leligdowicz, B.Sc. *.

INTRODUCTION

Over 50 million people are infected with the Human Immunodeficiency Virus (HIV) worldwide (1). Caring for the millions of people living with HIV and AIDS is essential. Over 40 therapies are approved by the Food and Drug Administration and as a result, the death rate has fallen dramatically across Europe and the United States. However, AIDS is increasingly a disease of the poor and medicines are where the problem is not, while the problem is where the medicines are not (2). Terminating the HIV epidemic is of utmost significance especially to developing countries, which account for over 95% of new HIV infections (1). Therefore, controlling the virus is important not only to the state of world health, but also to global development.

Glycobiology, the study of carbohydrates, is an interesting topic to medical research. Critical biological processes, including regulation of the growth and mobility of cells, immune responses, and responses of cells to hormones and growth factors, all depend on carbohydrates. In addition, viruses, including HIV, use cell-surface carbohydrates to get into cells and initiate infections. The aim of this review is to demonstrate how the expanding research in glycobiology can be applied to understanding the structure of HIV, as well as developing potential treatment approaches.

ESSENTIALS OF GLYCOBIOLOGY

One of the post-translational modifications that occurs on proteins after they have begun to be translated in the endoplasmic reticulum (ER) is N-linked addition of carbohydrate chains, resulting in the formation of glycoproteins (figure 1)

Glycoproteins contain one or more carbohydrate residues. The carbohydrate moieties of glycoproteins help to determine the tertiary structure of a protein. In addition, they may serve as biological labels, marking the proteins for different fates. This is possible because the number of possible permutations and combinations of monosaccharide types and glycosidic linkages in an oligosaccharide is astronomical (4). Each oligosaccharide can therefore present a different face that can be recognized by not only specific enzymes but more importantly for HIV infection, by specific receptors.

OVERVIEW OF HIV INFECTION

The human immunodeficiency virus can be classified as types one (HIV-1) and two (HIV-2). HIV-1 is older, more infectious, and accounts for the vast majority of HIV worldwide while HIV-2 is present mainly in Western Africa (5). Since HIV -1 accounts for the majority of the infections worldwide and since it is 3.55-fold more infectious than HIV-2, it will be the virus type of focus of this review.

HIV-1 is the causative agent for the acquired immunodeficiency syndrome, or AIDS. Infection begins when HIV interacts with a host cell surface membrane protein (CD4) via a viral envelope glycoprotein. The glycoprotein is an oligomer of extracellular (gp120) and transmembrane (gp41) glycoproteins (6). The gp120 is responsible for virion binding to CD4 receptors of host cells, whereas gp41 mediates fusion of the HIV virus and the host cell membrane. The fusion of membranes allows the viral

* To whom correspondence should be addressed: 418 Ave des Pins Ouest, Apt 34, Montreal, Quebec H2W 1S2, Canada.
E-mail address: alelig@po-box.mcgill.ca.

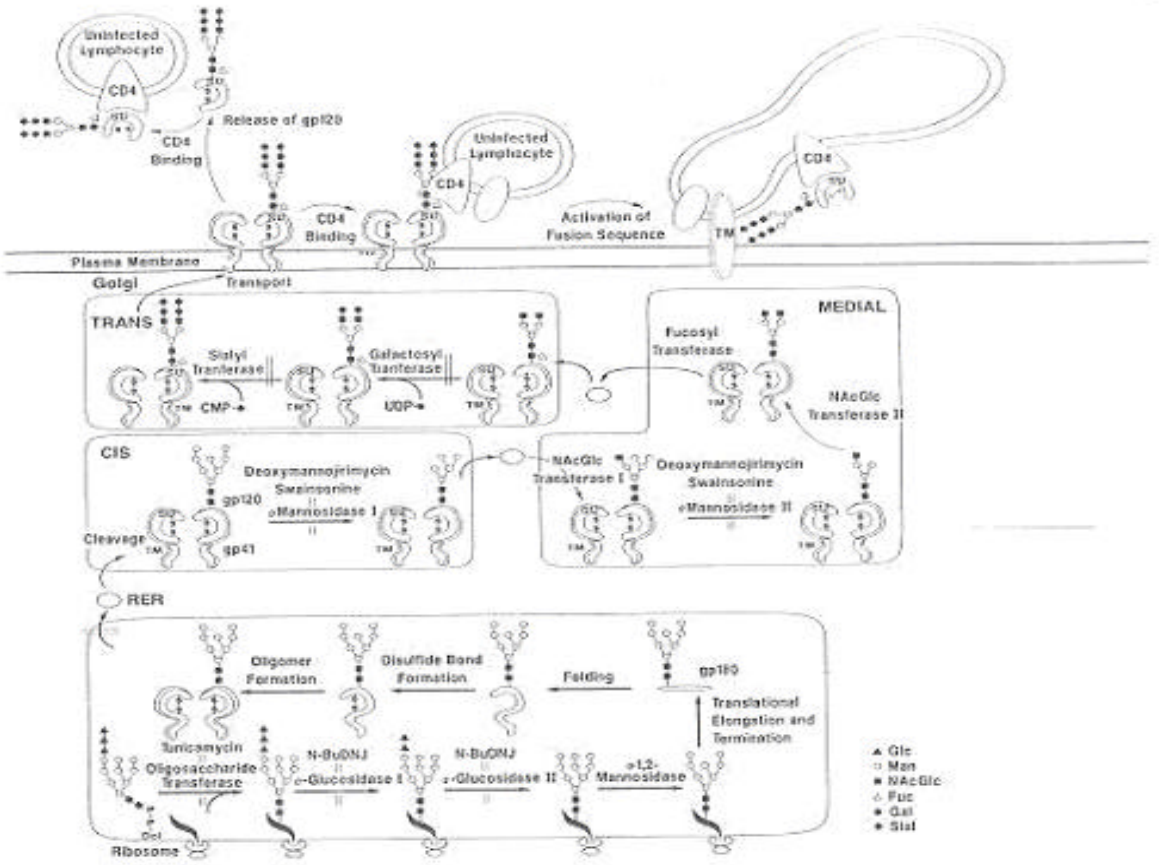


Figure 1. Steps in the formation of N-linked oligosaccharides in the lumen of the rough endoplasmic reticulum ER and the various compartments of the Golgi complex. The attachment of oligosaccharides is via an N-glycosidic bond to the side chain of an asparagine residue on the protein. The asparagine is part of the sequon Asn-X-Thr/Ser. The sequon is recognized by an enzyme that transfers the oligosaccharide from dolichol to Asn in the peptide. The first seven sugars (2 GlcNAc's and 5 Mannoses), are transferred one at a time to the dolichol-PP on the cytosolic side of ER. The oligosaccharide then flips across the ER membrane and the rest of the monosaccharides are added to the glycan on the luminal side of the membrane. After the oligosaccharide is assembled, it is transferred to an asparagnine residue on the nascent peptide. Once the glycan is attached, it undergoes removal steps mediated by glucosidases (glucose cleaving enzymes) that take off the terminal glucose residues. The protein is then shipped to the Golgi complex, where most of the terminal mannose units are removed by mannosidases and a variety of sugars.
Various Glycosylation pathway inhibitors are listed above the enzymes associated with the steps in the pathway (3).

particle to enter the cell by receptor-mediated endocytosis. Once the virus is inside the cell, it releases its RNA genome and the enzyme reverse transcriptase (RT). RT reverse transcribes the viral RNA into DNA. The HIV DNA moves into the nucleus where it is inserted into the host cell's genome. New HIV virions are made by the host cell, which replicates and transcribes the HIV DNA, and translates the viral transcript as well. Newly formed virions are released from the host cell, ready to infect the next CD4+ host cell (T helper-cell).

The precursor to gp120 and gp41 is gp160. This glycoprotein is heavily glycosylated in the ER lumen, where disulfide formation also occurs. The gp160 is cleaved intracellularly by a host protease in the late Golgi compartment to yield the noncovalently linked gp120 and gp41 complexes (figure 2 a, b, c), which are then transported to the host plasma membrane. At the surface, the complexes can be used by the emerging virus when it buds out from the cell to equip the virus for infection of new CD4 cells. In addition, the gp120-gp41 complex in the host plasma membrane can bind a CD4 receptor on an uninfected cell, resulting in cell fusion (6).

HIV GLYCOPROTEIN STRUCTURE

The use of mutated recombinant HIV glycoproteins to study their structure

The structure of the HIV envelope glycoproteins (gp120 and gp41) is important for developing drugs that can function to eradicate the virus. An understanding of the chemical nature and the processing of the oligosaccharide chains attached to the envelope proteins, and the identification of the sites where post-translational modification occurs on the gp160 precursor glycoprotein can serve as an invaluable tool

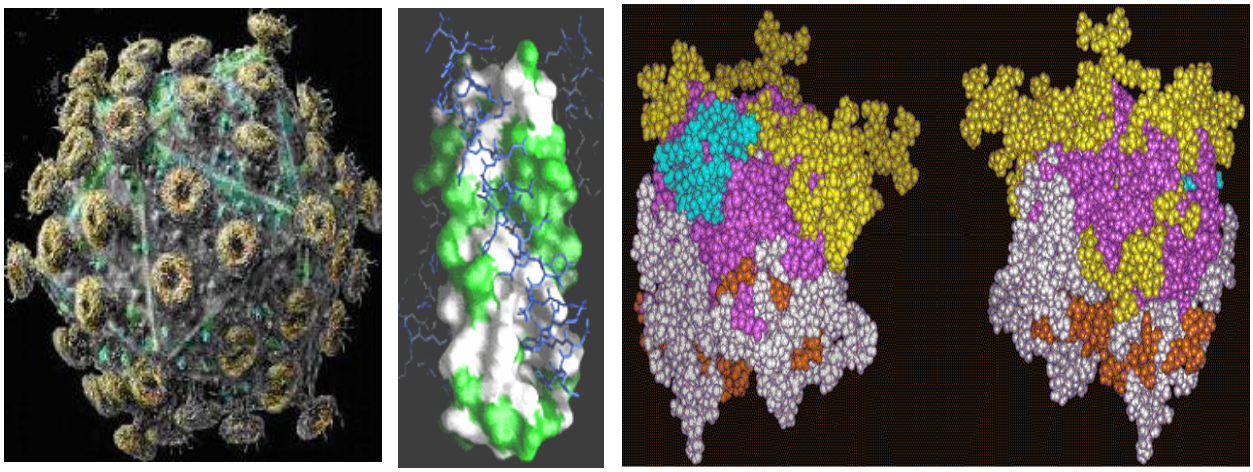


Figure 2. Structure of HIV
a) Structural view of the viral envelope (0.1 µm in diameter). Embedded in the envelope is a protein that consists of an outer protruding cap (gp120) and a stem (gp41) (10) . b) Close up view of the structure of gp41(11). c) X-ray crystallography image of gp120. Here, two views of the gp120 show the total topography of the glycoprotein. In white and yellow are the glycosylation sites(12).

for terminating the spread of the virus within the host.

Glycosylation of gp160 is extensive. The gp120 portion of gp160 has approximately 24 potential N-linked glycosylation sites (Asn-X-Thr/Ser) (7) and the weight of the glycans make up over 50% of the molecular mass of the gp120 glycoprotein(6, 8) (figure 2). In contrast, the gp41 ectodomain is poorly glycosylated, with only four to five potential sites for N-linked glycosylation (6). No evidence for O-linked glycosylation of the envelope proteins has been observed (9).

To study N-linked oligosaccharide structures on the HIV gp120, Mizouchi et al. used chronically HIV-infected lymphoblastoid (H9) cells and sequenced the glycans after they were released from proteins by hydrazinolosis. The conclusion from these studies is that the number of possible oligosaccharide structures present on gp120 outnumbers the 24 potential glycosylation sites present on the glycoprotein. Therefore, alternative structures occur on some of the glycosylation sites and numerous glycosylation variants of the gp120 are produced even in one cell line (13). The varied glycosylation is beneficial for the virus because it allows the virus to escape the immune system's response to the original invading virus by always appearing to the immune system as a new pathogen.

The importance of glycosylation sites in the infectivity of the virus has been investigated in both gp120 and gp41. This has been achieved by mutating the DNA sequences that code for the portion of glycoprotein important in its interaction with CD4 molecules on host cells. The results indicate that multiple mutations of glycosylation sites are required to alter the function of the gp120/gp41 complex (6). For example, the complete removal of glycan clusters from gp41 accomplished by altering the asparagines of the N-linked glycosylation sequons in positions 621, 630, and 642 to serine residues, abrogated the viral ability to carry out fusion completely (6). Mutated gp41 co-migrated with deglycosylated forms of wild-type gp41 on electrophoretic gels (14), suggesting that the deletion of the glycosylation sites contributes to the lack of glycosylation of the mutated proteins. These studies prove that for a successful HIV infection, the viral envelope proteins must be properly glycosylated so that the folding of the glycoproteins is correct and so that gp120 can interact with a CD4 host membrane receptor and allow the subsequent entry of the virus into the host cell. If gp41 is incompletely glycosylated, membrane fusion between the viral envelope and the host plasma membrane cannot occur (14). This finding may be useful in the construction of drugs that focus on the alteration of the oligosaccharide part of the HIV glycoprotein structure during its processing, decreasing the number of HIV particles that are capable of infecting CD4+ cells.

Additional studies involving mutated forms of the gp41 portion of gp160 have demonstrated that the transport of mutated forms of gp160 from the cis to the medial Golgi is slow and the transport to the trans Golgi is impaired. Therefore, since cleavage occurs in the trans part of the Golgi, cleavage reactions were severely impaired in the mutated gp160 at the gp120-gp41 junction. Gp160 lacking gp41 carbohydrates demonstrated that proteins lacking glycans are arrested in the Golgi following biosynthesis. Therefore, the glycan components of the glycoprotein are important in

the intracellular transport and processing of the gp160 (5).

N-linked carbohydrates within the epitope portion of the HIV gp120 are important in the recognition of the HIV CD4 receptor. Support for this comes from a study by Botarelli et al. who immunized T-cells with a recombinant non-glycosylated gp120. When the immunized cells were presented with a wild-type glycosylated form of gp120, the cells' T-cell receptor did not recognize the glycosylated gp120 and did not develop a rapid immune response to it. To investigate the reason for the slow recognition of glycosylated glycoprotein, the epitope portion of the gp120 bound to the T-cell was mapped. It was found that the epitope region of gp120 contained two asparagines residues which are sites for N-linked glycosylation (8). Therefore, the carbohydrates on the wild-type gp120 prevented epitope recognition by the immunized cells that had a T-cell receptor that was specific for the glycosylated form of gp120 used in the immunization.

The use of oligosaccharide processing inhibitors to study HIV glycoprotein structure

Previous studies have relied on utilizing glycobiology tools to better understand the mechanism of HIV infectivity and the source of its immunogenicity. For example, N-linked glycosylation processing inhibitors can be used for studying the importance of protein glycosylation in HIV pathogenesis, replication, target-cell infectivity, and syncytium formation (15). The effect of oligosaccharide processing inhibitors can be determined by electrophoretic mobility. Proteins with the lowest molecular weight travel the furthest on the electrophoretic gel. In studies where gp120 is processed in the presence of glucosidase inhibitors (castanospermine or 1-deoxynojirimycin), there is a decrease in gp120 mobility. However, when mannosidase inhibitors (1-deoxymannojirimycin or swainsonine) are present, glycoprotein mobility is increased (15). These results suggest that the final glycan products found on gp120 are smaller that the glycan products seen when glucosidase inhibitors are added and bigger than high mannose chains that are seen when mannosidase inhibitors are introduced into the media.

Inhibitors of oligosaccharide processing can also be used to study the steps in the processing of envelope glycoproteins in cells. For example, treating chronically HIV-infected cells with tunicamycin (an inhibitor that completely abrogates the addition of glycans to N-linked glcosylation sites of the protein) severely inhibits the glycosylation of envelope proteins. Treatment with deoxynojirimycin, an inhibitor of glucosidase I in the rough ER, inhibits proteolytic cleavage of gp160 (9, 15). Inhibitors of mannosidase I and mannosidase II (deoxymannojirimycin and swainsonine respectively) allow for the processing of the gp160 to gp120 and gp41. However, the virions made by cells treated with mannosidase inhibitors are significantly less infectious than virions synthesized by untreated cells. This suggests that proteolytic cleavage of gp160 takes place in infected cells when the glycoprotein has mannose-rich oligosaccharide structures. However, for the release of infections virions, the trimming of glucose residues and the primary trimming of mannose are both necessary (15). Therefore, oligosaccharide processing inhibitors can be used to decrease the infectivity of new HIV virions since the viral particles would contain improperly glycosylated glycoproteins.

HIV THERAPY BASED ON CURRENT KNOWLEDGE OF GLYCOBIOLOGY

HIV is a very complicated pathogen due to its ability to mutate at a fast rate. It introduces difficulty for the immune system, which continually sees the virus as foreign, posing challenges to treatment. Research in the field of HIV has contributed to a wealth of therapeutic options. Current treatments can be classified into the following classes: (i) nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs) (i.e. zidovudine (AZT), (ii) non-nucleoside reverse transcriptase inhibitors (NNRTIs) (i.e. nevirapine); and (iii) protease inhibitors (PIs) (i.e. saquinavir). Other steps in the HIV replicative cycle that are potential targets for chemotherapeutic intervention include viral adsorption, viral entry, virus-cell fusion, proviral DNA integration, and viral mRNA transcription (16). Additionally, the expanding field of glycobiology proposes additional treatment options.

HIV drug therapy based on inhibitors of oligosaccharide processing

Application of carbohydrate analogues such as tunicamycin (inhibitor of the synthesis of the dolichol-linked oligosaccharide precursor), castanospermine and 1-deoxynojirimycin (DMN, glucosidase inhibitor), as well as 1-deoxymannojirimycin (mannosidase I inhibitor), has been shown to attenuate the HIV-1 infectivity invitro. These inhibitors oligosaccharide processing are ineffective in preventing the formation of a cytoplasmic continuity (syncytia) between two cells when gp120/gp41 complexes are targeted to the infected cell's plasma membrane and bind to CD4 receptors of neighboring CD4+ cells (15). However, N-glycosylation is critical in HIV pathogenesis at the level of viral binding and fusion to uninfected CD4+ cells.

The gp120/gp41 HIV envelope complex must be

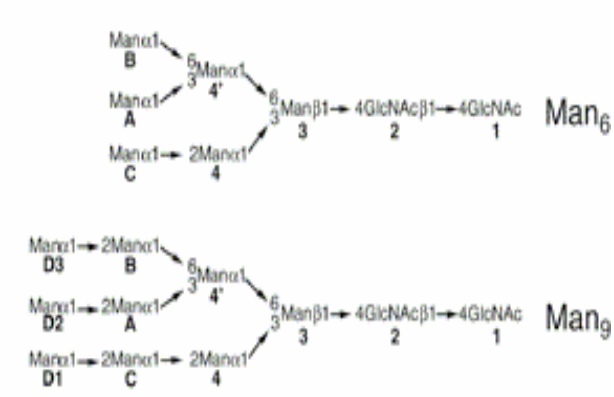


Figure 3. Formulas of N-linked oligomannose molecules including Man(6) and Man(9). Isomers of Man(7) and Man(8) can be deduced by adding mannose termini to the Man(6) structure (18).

folded properly for it to be pathogenic. Proper folding is mediated by oligosaccharides on the gp120/gp41 complex. Therefore, glycosylation inhibitors can be applied to alter the fusogenicity of the gp120/gp41 complex since improperly folded proteins cannot induce fusion. For example, DNM inhibits the cellular a-glucosidase I-II activity, blocking the trimming of the glycan precursor at the level of the ER (i.e. the cleavage of three glucose residues from Glc(3)Man(9)GlcNAc(2) precursor glycan, see figure 3). In the presence of DNM, glycoproteins with abnormally glycosylated oligomannosidic moieties are produced. Treatment of HIV-infected lymphocyte cultures with DNM inhibited the spread of the virus (16). This was possible because although the gp120/pg41 complex synthesized in the presence of DNM could bind CD4 receptors on human lymphocytes, it was not able to induce membrane fusion. N-butyl DNM, an alkyl substituted variant of DNM, has shown antiviral effects at doses with no reported cell cytotoxicity (3). Phase I clinical trials on n-butyl DNM in twenty-nine patients, however, showed no significant trends in CD4+ cell counts, and Grade II elevations in liver function, leucopenia, and neutropenia were observed in some patients, calling the study to a halt before the dose-escalading tolerance trial could be completed (17).

HIV drug therapy based on oligosaccharide structure

If the glycan structure composition of the HIV glycoprotein gp120 can be determined, the interaction between gp120 and CD4+ cell can become a potentially useful target in HIV therapy. To study the gp120 epitopes involved in interaction with the CD4 receptor, three N-linked glycosylation sites on the gp120 in the vicinity of the epitope recognized by the host T-cell were removed (7). Mutant gp120 with one, two, and three N-linked glycans missing were analyzed. The conclusions made were that two of the three glycosylation sites were heterogeneous in the structural composition of glycans. However, the site located next to the T-cell epitope had a large, high mannose structure with more than 11 mannose units and it covered a large part of the surface of gp120 (7). The finding that high mannose glycan structure present in the epitope of the gp120 is recognized by the T-cell has allowed for the creation of cyanovirin-N (CV-N), a drug that inactivates diverse strains of HIV at the level of cell fusion by virtue of its interaction with specific N-linked oligosaccharides on gp120.

CV-N is a 11kDa monomeric protein derived from cyanobacteria. CV-N binds to glycosylated gp120 but not to unglycosylated gp120. When the ligand which binds to CV-N was studied by examining the binding of gp120 to a CV-N affinity column, only Man-8 and Man-9 glycoforms (figure 3) were preferentially retained on the affinity column. Studies have found that adding free Man-8 and Man-9 oligosaccharides partially inhibited binding of CV-N to gp120 (2). When a mixture of carbohydrates that structurally represent N-linked carbohydrates found on the gp120 was added, it was found that CV-N specifically recognizes with nanomolar affinity Man(9)GlcNAc(2) and Man(8)GlcNAc(2) (18). Viruses sensitive to CV-N exhibit an abundant exposure of high mannose oligosaccharides on their surfaces (figure 3) (19). All of the above findings suggest that the part of gp120 which is recognized by CV-N is the high mannose glycan portion of the glycosylated gp120.

CV-N is unlike other lectins that have been reported to have anti-HIV activity mediated through interactions with gp120. Classical carbohydrate-interacting lectins, such as concanavalin A and wheat-germ agglutinin, associate with gp120 in a monosaccharide-specific manner and are inhibited by the presence of exogenous monosaccharides. However, CV-N/gp120 interaction is not inhibited by high concentrations (i.e. 10,000-fold excess) of monosaccharides (19). This observation suggests that CV-N/gp120 interactions are defined by a more complex oligosaccharide-specific binding. In addition, the carbohydrate structures recognized by CV-N are rare in normal human tissue. Therefore, CV-N is a potential oligosaccharide-specific therapeutic agent for the treatment of HIV as well as other pathogens with high mannose ligands on their outer surface.

Besides CV-N, other drugs have also been designed based on the principle that high mannose chains are essential for HIV infection. For example, Pradmicin A, an antifungal antibiotic isolated from Actnomdura hibisca, and its derivative, BMY-28864, have the ability to inhibit HIV infection in vitro (20). Similar to the CV-N studies, the inhibitory effects of these drugs were suppressed by the addition of high mannose type

oligosaccharides. Therefore, targeting oligosaccharide chains of the envelope glycoprotein is a possible way to block HIV infection.

Therapy based on oligosaccharide structure to construct a potent HIV vaccine

An effective way of combating HIV is via vaccination. To construct a successful vaccine, the HIV gp120 antigen needs to be recognized by the helper T-cell. In order to generate vaccines that will induce a strong immune response, the HIV gp120 antigen needs to be made more foreign. In one vaccination study, gp120 was made more immunogenic by neuraminidase treatment, which removed terminal salicylic acids from the carbohydrate side-chains of the glycoprotein, producing an asialoglycoprotein (21). Neuraminidase treatment exposed terminal galactose residues so that they could be recognized by galactose receptors on antigen presenting cells. Evidence for the use of galactose receptors became apparent when galactose is added to cultures of antigen presenting cells. Extrinsic galactose, when added, competes with asialo-glycoproteins for binding to galactose receptors. Thus, the antigenicity of the HIV glycoprotein gp120 can be enhanced by exposing galactose residues on the gp120 since these glycoproteins can be internalized by antigen presenting cells containing galactose receptor. Uptake of the proteins with terminal galactose units results in a quicker immune response because asialoglycoproteins are more quickly recognized as foreign (4).

When constructing an HIV vaccine, it is therefore worthy to explore glycobiology to produce a vaccine with a more immunogenic antigen. If a more antigenic form of gp120 could be administered to a patient, the patient may be able to mount a stronger immune response against the HIV.

Inhibition of HIV gp120 binding to CD4 molecules by natural glycosaminoglycans

It is a fact that HIV is spread via body fluids. Past studies of HIV inhibitors have suggested that the presence of inhibitory macromolecules in the body fluids that do not transmit the virus could be responsible for the inability for these fluids to function as a vehicle for HIV transmission. Other body fluids that serve as a route of HIV transmission, such as breast milk, have been found to contain compounds that could limit the rate of postnatal vertical transmission for HIV in breast-fed infants of HIV-infected mothers. The proposed reason for the lack of viral transmission by some body fluids is the presence of glycosaminoglycans (22).

Previous studies have proven that non-antibody glycoconjugates from human milk can inhibit the binding of many pathogenic microorganisms. According to Newburg et al, glycosaminoglycans found in human milk represent a novel class of naturally occurring molecules that are capable of inhibiting viral binding to host CD4 receptors (22).

Glycosaminoglycans are heteropolysaccharides; they are linear polymers made of repeating disaccharide units. One of the monosaccharides is always N-acetylglucosamine or N-acetylgalactosamine, and the other is a uronic acid such as glucuronic acid. One or more of the hydroxy groups of the amino sugars can be sulfonated (4), producing glycosaminoglycans such as dermatan sulfate, heparan sulfate, or chondroitin sulfate.

Sulfonated molecules, including sulfonated glycosaminoglycans (GAG), have been shown to inhibit the binding and replication of HIV. The classes of GAG found in human milk include heparin, heparan sulfate, chondroitin sulfate, and dermatan sulfate .

To determine the types of glycosaminoglycans that are most important in the inhibition of HIV replication, Newburg et al analyzed milk from thirty women. To test for the inhibitory nature of each glycosaminoglycan from human milk, isolated GAGs, CD4 and gp120 molecules were mixed and allowed to interact. Glycosaminoglycan cleaving enzymes (lyases) were then added to cleave the linear glycosaminoglycan polymeric backbone and the reaction mixtures were tested for the loss of inhibitory action. The greatest loss in the inhibition of the gp120 and CD4 interaction was seen when chondroitinase AC was added (22). This suggests that the specific glycosaminoglycan component of human milk responsible for the inhibition of HIV binding and replication is chondroitin sulfate.

Further research on sulfated glycosaminoglycans could provide valuable treatment methodologies to terminate the HIV epidemic. Based on the studies of sulfated glycosaminoglycans, synthetic sulfonated polymers that share the gross features of GAGs could be synthesized such that they contain anti-HIV activity.

CONCLUSION

New antiretroviral therapies prolong the lifespan of people infected with HIV. Correspondingly, the number of AIDS-related deaths has declined. Despite the availability of drug therapies, HIV is still a challenge. Glycobiology is a field with foreseeable therapeutic alternatives for HIV. Glycobiology-related treatments such as inhibitors of oligosaccharide processing, treatments based on blocking the reactive sites of gp120, vaccinations, and natural methods using glycosaminoglycans to combat the deadly virus, further add to the list of potential antiretroviral treatments. Combining these treatments with current antiretroviral methods could result in synergistic anti-HIV effects.

However, further testing of the glycobiology-related therapies is needed to determine the non-specific effects on other cellular processes and before the therapies can be made available for the HIV-affected population.

REFERENCES

1. International AIDS Vaccine Initiative. http://iavi.org ; 2002.
2. Bloom, DE. Something to be done: treating HIV/AIDS. River Path Associates. Science 288(5474):2171-3; 2000.
3. Ratner L. Glucosidase Inhibitors for Treatment of HIV-1 Infection. Aids Research and Human Retroviruses 8:165-173; 1992.
4. Lehninger Albert L, Nelson David L, Cox Michael M. Principles of Biochemistry, 2nd edition. New York, NY: Worth Publishers, 1993.
5. Gilbert PB, McKeague IW, Eisen G, Mullins C, Gueye-NDiaye A, Mboup S, Kanke PJ. Comparison of HIV-1 and HIV-2 infectivity from a prospective cohort study in Senegal. Statist. Med 22:573-593; 2003.
6. Pal R, Hoke, GM, Sarngadharan MG. Role of gp41 Glycosylation Sites in the Biological Activity of Human Immunodeficiency Virus Type 1 Envelope Glycoprotein. Virology 242:338-345; 1998.
7. Biller M, Bolmstedt A, Hemming A, Olofsson S. Simplified procedure for fractionation and structural characterization of complex mixtures of N-linked glycans, released from HIV-1 gp120 and other highly glycosylated viral proteins. J Virol Methods 76:87-100; 1998.
8. Botarelli P, Houlden BA, Haigwood NL, Servis C, Montagna D, Abrignani S. N-glycosylation of HIV-gp120 may constrain recognition by T lymphocytes. J Immunol 147:3128-3132; 1991.
9. Kozarsky K, Penman M, Basiripour L, Haseltine W, Sodroski J, Krieger M. Glycosylation and processing of the human immunodeficiency virus type 1 envelope protein. J Acquir Immune Defic Syndr 2:163-169; 1989.
10. AVERT, AIDS Virus Education and Research Trust. www.avert.org/virus.htm ; 2002.
11. Howard Hughes Medical Institute. Key HIV Protein Structure Solved. http://www.hhmi.org/news/keyhiv.html ; 1997.
12. Zhu X, Borchers C, Bienstock RJ, Tomer KB. Mass Spectrometric Characterization of the Glycosylation Pattern of HIV-gp120 Expressed in CHO Cells. Biochemistry 39:11195;11204; 2000.
13. Mizuochi T, Matthews TJ, Kato M., Hamako J, Titani K, Solomon J, Feizi T. Diversity of oligosaccharide structures on the envelope glycoprotein gp120 of human immunodeficiency virus 1 from the lymphoblastoid cell line H9. Presence of complex-type oligosaccharides with bisecting N-acetylglucosamine residues. J Biol Chem 265:8519-8524; 1990.
14. Fenouillet E, Jones I, Powell B, Schmitt D, Kieny MP, Gluckman JC. Functional role of the glycan cluster of the human immunodeficiency virus type 1 transmembrane glycoprotein (gp41) ectodomain. J Viro, 67:150-160; 1993.
15. Pal R, Hoke GM, Sarngadharan MG. The role of oligosaccharides in the processing and maturation of envelope glycoproteins of human immunodeficiency virus type 1. Proc Natl Acad Sci USA 86:3384-3388; 1989.
16. De Clercq E. New developments in anti-HIV chemotherapy. Curr Med Chem 8:1543-72; 2001.
17. Tierney M, Pottage J, Kessler H, et al. The tolerability and pharmacokinetics of N-butyl-deoxynojirimycin in patients with advanced HIV disease (ACTG 100). The AIDS Clinical Trials Group (ACTG) of the National Institute of Allergy and Infectious Diseases. J Acquir Immune Defic Syndr Hum Retrovirol 10:549-553; 1995.
18. Bewley CA, Otero-Quintero S. The potent anti-HIV protein cyanovirin-N contains two novel carbohydrate binding sites that selectively bind to Man(8) D1D3 and Man(9) with nanomolar affinity: implications for binding to the HIV envelope protein gp120. J Am Chem Soc 123:38923-902; 2001.
19. Bolmstedt AJ, O'Keefe BR, Shenoy SR McMahon JB, Boyd MR. Cyanovirin-N defines a new class of antiviral agent targeting N-linked, high-mannose glycans in an oligosaccharide-specific manner. Mol Pharmacol 59:949-954; 2001.
20. Mizuochi T, Nakata M. New anti-HIV drug which binds the oligosaccharides of HIV envelope glycoprotein. Nippon Rinsho 53:2340-2349;1995.
21. Manca F. Galactose receptors and presentation of HIV envelope glycoprotein to specific human T cells. J Immunol 148:2278-2282; 1992.
22. Newburg DS, Linhardt RJ, Ampofo SA, Yolken RH Human Milk Glycosaminoglycans Inhibit HIV Glycoprotein gp120 Binding to Its Host Cell CD4 Receptor. J Ntr 125:419-423; 1995.
23. Medical microbiology, Glycosaminoglycans. http://www.indstate.edu/thcme/mwking/glycans.html ; 2001.

Aleksandra M. Leligdowicz is a second year medical student at the McGill University Faculty of Medicine. In 2000, she completed a BSc degree in molecular biology at the University of Manitoba and Deakin University, Australia. In October 2004, she will begin research in Immunology and Epidimiology of HIV-2 at the Institute of Molecular Medicine at Oxford University as a Rhodes Scholar.

REVIEW ARTICLE

The Vegetative State: A Review of Etiology and Prognostic Factors

Rachel Davison*

ABSTRACT This paper reviews the research investigating the vegetative state (VS) in terms of its aetiology and prognostic factors that may be indicative of the outcome for patients in the VS. The VS is a relatively rare syndrome that still causes confusion for treating clinicians. In short, the VS is a clinical condition of unawareness of self and environment but with retained wakefulness. Until relatively recently there were no universally accepted diagnostic criteria, which caused problems both in terms of diagnosing the patient and in determining the incidence of the VS. This paper examines the most relevant and up to date work in order to determine if there is a way of predicting whether the VS for any given patient will be persistent (i.e. recovery is still possible) or if it is permanent and further treatment is futile. Currently, the most accurately available method to predict the prognosis of a patient in the VS is through clinical assessment of the patient combined with knowledge of the aetiology and duration of the VS. More work is needed in order to allow for the prediction of the outcome of the VS with greater certainty.

INTRODUCTION

While the advent of cardiopulmonary resuscitation during the 1960s was a breakthrough for medical science, some survivors remained in a limbo state of 'waking unconsciousness'. It was Jennett and Plum (1), in 1972, who termed this condition the persistent vegetative state (PVS) and described the vegetative state (VS) as:

> The absence of any adaptive response to the external environment, the absence of any evidence of a functioning mind which is either receiving or projecting information in a patient who has long periods of wakefulness

In short, the VS is a clinical condition of unawareness of self and environment but with retained wakefulness. Efforts to predict the outcome of patients in a vegetative state began around this time due to the concern that large numbers of patients surviving in a VS would be costly and use resources that could be more effectively spent elsewhere (2).

The incidence of the VS is unknown, partly because of the rarity of the condition and partly because of the lack of accepted universal diagnostic criteria. Estimates range from 0.4-1.1/100,000 people throughout the world (3,4). Moreover, until 10 years ago, the VS was not a codable diagnosis in either the International Classification of Diseases or most health agencies. Studies at the time suggested that the prevalence in the United States alone was around 10,000-25,000 adults and 4000-10,000 children (3,4).

The VS causes distress to family and friends and creates difficulties for the doctors involved. To the non-medically trained, a patient in the VS may appear to be alive and functioning; for example they may seem to smile or turn to sound despite the fact that by definition the patient is not aware. This may cause conflict between the family of the patient and the doctors who are trying to explain what the VS is and the likely outcome. This may be further complicated by the fact

* To whom correspondence should be addressed:
Framlington Place University of Newcastle upon Tyne Medical School Newcastle upon Tyne England NE1 7RU
E-mail: R.E.Davison@ncl.ac.uk

that prediction of recovery is an inexact science. Doctors cannot give a definite answer and the family may prefer to simply watch and wait while they feel there is still hope. To provide the care that the patient needs, the doctor must have the ability to predict the best possible outcome and also to recognize when it would be more humane to withdraw medical interventions and let nature take its course. The ethical dilemmas regarding quality of life and best intentions are many and complex, and as such, are beyond the scope of this review. This paper reviews the research investigating the vegetative state (VS), in terms of its etiology and prognostic factors that may be indicative of the outcome for patients in the VS. Furthermore, this review identifies the methods by which prediction of prognosis may be possible.

METHODS

A literature search was conducted in Medline (from 1993) and EMBASE (from 1980) to identify suitable papers. The databases limited the choice of date selection; the selected dates were chosen to identify only the most up to date work. The keywords used were PVS, VS, (persistent) vegetative state, (permanent) vegetative state, children and (persistent, permanent) vegetative state, coma or life support care, combined with prognosis or prediction. The aim in the use of the keywords was to detect all relevant articles in the given time period. In all, about one hundred articles were identified but not all of the papers were applicable to the aims of this review. The reference lists of relevant articles were scrutinised to detect any additional studies that had not been already identified. The articles most applicable to the subject were selected and the information within them collated, with careful attention being paid to the methods of the systematic reviews and critical analysis of the original studies that warranted inclusion.

CONSCIOUSNESS AND THE VEGETATIVE STATE

There are two components to consciousness (3):

1. Wakefulness

Clinical work indicates that the midline structures in the upper pons, midbrain and thalamus (the reticular activating system or RAS) are necessary for wakefulness. They are activated by arousal and influence the cerebral cortex directly. Wakefulness is essentially not being asleep; therefore it comprises acts that one would not do while asleep such as opening one's eyes and looking around.

2. Awareness

The cerebral cortex and its projections to the major subcortical nuclei are considered to be the root of awareness with its content being the mass of information that it processes from the external environment. Awareness comprises behaviours that indicate that a person comprehends the outside environment (e.g.: communication and understanding).

As one of the main links between the two component areas, the thalamus is crucial to the preservation of consciousness (3,5). Unconsciousness therefore implies global (or total) unawareness. Awareness requires wakefulness, but wakefulness can be present without awareness. In the comatose state both awareness and wakefulness are lacking, while in the VS wakefulness is preserved and awareness is not (3).

If the VS lasts for a month, it is termed continuing or persistent VS (3,6). The consensus is that the terminology changes to that of permanent VS when the condition is deemed to be irreversible, no recovery seems possible and further treatment is considered futile. This decision is usually taken once a year has elapsed in traumatic aetiologies and after three months in non-traumatic cases (4-6). However, as with all clinical judgements, it is based on probabilities. While persistent VS is a diagnosis, permanent VS is a prognosis. In practise, the terms are commonly used interchangeably and the acronym PVS is used for both conditions (3,5-7). Therefore, the acronym VS will be used in this review, and distinguished where necessary if specific reference is being made to a persistent or permanent state. Recovery from the VS is classified by the Glasgow Outcome Scale (GOS, see Table 1), with a good recovery indicated by a GOS score of 4 or 5.

Table 1. The Glasgow Outcome Scale (GOS). (4).

Level	Term	Definition
1	Dead	No life
2	Vegetative state	Unaware of self and envirion
3	Severe disability	Uable to live independently
4	Moderate disability	Able to live independently
5	Mild disability	Able to return to work/school

Diagnosing the VS

Before a diagnosis of the VS can be made, an established cause must be found and all reversible factors that may be contributing (e.g. metabolic disturbances, sedatives, anaesthetics or neuromuscular blocking drugs) eliminated (5).

The Multi-Society Task Force (MSTF) on persistent VS defined the following diagnostic criteria, which are widely acknowledged by the medical community (3):

1. No evidence of awareness of self or environment and an inability to interact with others.
2. No evidence of sustained, reproducible, purposeful or voluntary behavioural responses to visual, auditory,

tactile or noxious stimuli.
3. No evidence of language comprehension or expression.
4. Intermittent wakefulness manifested by the presence of sleep-wake cycles.
5. Sufficiently preserved hypothalamic and brainstem autonomic functions to permit survival with medical and nursing care.
6. Bowel and bladder incontinence.
7. Variably preserved cranial nerve reflexes (pupillary, oculocephalic, corneal, vestibulo-ocular and gag) and spinal reflexes.

Patients in the VS are usually not immobile. There may be apparent semi-coordinated movements such as scratching, moving hands towards noxious stimuli (such as during mouth care) and reflex grasping. There is flexor withdrawal after a delay; inflicting a painful stimulus such as pressing the supraorbital ridge causes a stereotyped flexing of limbs as during assessment of Glasgow Coma Scale status (GCS, see Table 2). Movements are slow, dystonic and obviously abnormal. Neck movements may provoke reflex postural alterations. There may be chewing and grinding of teeth, and food and liquid placed into mouth may be swallowed. Patients in the VS may retain the response of turning their head or moving their eyes to sound, but eye fixation and tracking is not demonstrated. They may smile or appear to shed tears, and may grunt or groan (vocalise) but never speak (verbalise) (3,5-7). Although this may be disturbing to family and carers, it should be remembered that by definition patients in the VS have no awareness, and therefore it is the opinion of experts on the subject that they cannot feel any pain (4-7).

Table 2. The Glasgow Coma Scale (GCS).

Eye Opening (eGCS)	Motor Response (mGCS)	Verbal Response (vGCS)
1. No response	1. No response	1. No response
2. Open to pain	2. Abnormal Extension	2. Incomprehensible
3. Open to verbal command	3. Abnormal Extension	3. Inappropriate
4. Open spontaneously	4. Withdrawal	4. Confused
	5. Localises to pain	5. Fully orientated
	6. Obeys commands	

Diagnostic problems

Determining cognitive awareness in another person can only ever be an educated guess as there are no tests that can confirm the presence or absence of inner awareness (3,6-8). Repeated assessment is therefore essential, especially if there is some doubt over whether the behavioural patterns necessary for diagnosis of the VS are present.

The diagnosis of the VS is difficult to make in infants younger than three months because they have a limited capacity to show higher cognitive functions; the differentiation between voluntary and involuntary responses may also be unreliable until this age. The concept of the VS cannot be applied to preterm infants because of developmental immaturity and the lack of consistently recognisable sleep-wake cycles. The exception is infants born with severe developmental malformations such as anencephaly and hydranencephaly where there is minimal or no cerebral cortex and therefore no awareness. These infants are categorised as being in a VS congenitally (3,8).

PATHOPHYSIOLOGY AND ETIOLOGY OF THE VS

The VS is largely characterised by a functioning brainstem with no input from the cerebral hemispheres due to either disconnection or damage of two main types (3):

1. *Acute*, e.g. head injury, hypoxic-ischaemic damage following cardiopulmonary arrest, metabolic disturbances.
2. *Chronic*, e.g. degenerative processes such as Alzheimer's disease, congenital defects such as anencephaly.

The VS can be caused by a vast array of conditions; with any insult to the body that causes damage to the cerebral cortex being a potential cause of the VS. Three main patterns of brain pathology are seen at autopsy (3,9):

1.*Diffuse Axonal Injury* (DAI), which is extensive subcortical axonal injury that virtually isolates the cortex from other parts of the brain. It is most commonly due to the shearing forces in trauma or sometimes to hypoxic-ischaemic insults and is the most common pathological feature seen.
2.*Extensive laminar necrosis* is due to acute global cerebral ischaemia or hypoxia. Multifocal or diffuse necrosis is seen with almost invariable involvement of the hippocampus, hypothalamus and brainstem.
3.*Relatively selective thalamic necrosis* is an uncommon observation that may follow acute global ischaemia. Specific anatomical boundaries are not well described.

Mixtures of all three lesions are commonly seen with additional focal lesions depending on the precipitant of the insult. Acutely inflicted hypoxic-ischaemic insults and shearing forces are therefore shown to have a devastating impact on the brain, as these are the most frequently seen pathologies at autopsy. There have also

been reports of rare isolated lesions of the brainstem or hypothalamus alone causing the VS but these are not well studied (3).

The diagnosis of permanent VS is made by identification of cause, fulfilling diagnostic criteria and lasting for at least a set amount of time (three months for non-traumatic aetiologies, one year for traumatic) (6). Therefore, there are two aspects to the prediction of whether the VS will be permanent or whether recovery is possible - the etiology of brain insult and the duration of the VS to date.

1. Etiology of Brain Insult

Outcome of coma is directly related to its cause (2), which can be separated into two aetiologies: traumatic (e.g. road traffic accidents, falls) and non-traumatic (e.g. cardiac or pulmonary arrest, anoxic- ischaemic, metabolic).

Traumatic

Traumatic etiology of brain injury is the better of the two categories. The MSTF collated data from several similar studies of traumatic brain injury giving information on outcome for 434 patients (4). Recovery of consciousness varied with time, with 88% of those who recovered (46% of all patients studied) doing so within the first six months, and reaching 99.98% at one year. After this, recovery was rare. Only seven of the 434 (0.02%) recovered after this time, between one and three years after the injury. Five of them remained severely disabled, one was moderately disabled and the status of the seventh was undeterminable. Five of the seven were under 30 years, suggesting that age is another important confounding factor in the prediction of prognosis. Accordingly, those under 40 had a greater chance of recovering within three months without severe disability (4). Children in traumatic coma generally have a better prognosis than adults in a similar condition, although recovery of function is comparable (4,10). Data from several similar studies showed that 62% of children had regained consciousness at one year following injury, compared to 52% of adults (4,8).

Traumatic brain injury is also associated with a poor chance of a good functional recovery as described by the GOS (see Table 1). Using the collated data (4), from the 434 patients in a VS, of the 52% who had recovered consciousness by one year, 28% had severe disability, 17% moderate disability and only 7% had made a good recovery. Of those who made a good recovery over half showed signs of improvement within the first three months and all within six months. Those who recovered consciousness but remained disabled all began to show signs of improvement three to six months after the brain injury. This indicates that a later recovery was almost always associated with severe disability (4).

Non-traumatic

Non-traumatic etiology carries a much poorer prognosis than its traumatic counterpart. Collecting data on the outcome of 169 patients in a systematic analysis (4) showed 85% or more died within the first month or remained in a VS. Of the remaining 15% who recovered consciousness (11% of which were within the first three months, and the other 4% by six months), only one patient (0.6%) made a good recovery. One year after the injury, 32% remained in persistent VS and 53% had died. A larger study of 500 patients by Levy et al. (11) found that at one year, 73% were dead or in a VS and of the remaining 27%, 11% were severely disabled, 4% were moderately disabled and 12% (2.4% of the total) had made a good recovery. They agreed that most improvement occurred within the first month and that the longer the VS persisted, the smaller the chance of recovery.

There is very little evidence to suggest that there is a consistent relationship between age and prognosis in non-traumatic coma, mainly due to a lack of data (4,11). Shewmon (10) suggests that in children (under 16 years), the outcome is either full neurological recovery or remaining in a VS with very few outcomes in-between, with ratios ranging from 50:50 to 70:30 in favour of intact functioning. The commonest cause of non-traumatic coma in children is near drowning, and Shewmon (10) postulates that childrens' brains are more protected against anoxic-ischaemic damage due to their body temperature falling faster because of their smaller size. No evidence was found to confirm this, but it appears that children are less susceptible to anoxic or ischaemic injury and have a greater potential for neurological recovery than adults. As a result of this, the observation period in children is often allowed to be longer than the standard three months that adults are given before their VS is declared permanent (8).

Specific etiology

The underlying cause of the coma has been shown to relate to outcome in many studies (2,4,11). Metabolic and diffuse disorders carried a better prognosis than hypoxic-ischaemic causes (2,4,11). Cerebrovascular disease such as subarachnoid haemorrhage or stroke and other disorders causing structural brain damage carried the worst prognosis of all (2,4,11). The incidence of a VS was also observed to be higher following anoxic-ischaemic injury; for 20% of the patients studied this was the best outcome that they ever achieved (GOS 2) (2,11). Drug overdose also carried a favourable prognosis due to the reversibility of their effects, despite a bleak outlook at initial assessment that was considered to be due to depressive effects on the brainstem (2).

2. Depth and Duration of Coma

The longer the patient remains in a coma, the poorer their chance of recovery and the greater the chance that they will enter a VS (2,11). A week is often used as a cut-off point; by that time the chance of a moderate or good recovery is only 6-7% and almost half of those who are still unconscious will be in a VS (2,11). By convention, one month after the brain injury the patient is in a persistent VS (6) and the chance of recovery is small. However, studies by Andrews (12) and Dubroja and colleagues (13) have shown that recovery is possible beyond this time. In one case, recovery began three years after the initial insult (13). Whilst this shows that higher functioning can be regained beyond predictions, it should be noted that these studies only looked at small groups and that no one in them made a complete recovery. There have been other documented cases of very late recovery (4) but in all of them the level of function was far from independent. Such outcomes may be undesirable to some, but to others this may be an acceptable quality of life.

PROBLEMS WITH PROGNOSIS ACCURACY

Predicting prognosis in the VS is an inexact science and there are faults that are common to all the studies considered. The rarity of the condition itself (often resulting in studies with small sample sizes) coupled with the inability to conduct true prospective studies has led to many of the published papers lacking sufficient power to demonstrate their value. There are inevitable confounding factors in the studies, such as patients dying of non-neurological disease during the studies and the fact that in many studies no distinction is made between the VS and death, or the VS is combined with severe disability as a non-acceptable outcome. The studies are also limited by ethical considerations, as ideally patients would be kept alive indefinitely. Doctors are bound to act in their patient's best interests if the patient cannot express their own views. It can be argued, that keeping a patient in the VS simply for means of a scientific study is not in the patient's best interests and therefore is unethical. Short follow-up times are also commonplace. In addition, the studies face the problem of self-fulfilling prophecies [i.e. if a patient is predicted to have a poor outcome, it has been shown that therapies are often less aggressive and the next of kin are more likely to ask for withdrawal of medical support (2)].

Improvement on previous studies is difficult. Despite recognised problems with their methodology, it is not easy to correct them. We cannot alter the small sample sizes or the lack of prospective studies because of the rarity of the condition itself and because the ethical problems remain the same. This will most likely continue to be a problem unless a large scale, multicentre trial is organised with the cooperation of a large number of major neurological centres around the world.

Age

As has already been implied, the age of the patient may hold prognostic significance. Shewmon (10) states that children (under 16 years) have a better chance of recovering consciousness, although their functional recovery is often equivalent to that of adults. He asserts that the mortality rate from severe head injury declines with increasing age in childhood, reaching a trough at 14-15 years and then rising with age throughout adulthood. However, others have shown that outcomes worsen with rising age, even in childhood, which is possibly related to age specific differences in types of trauma (e.g. falls are more common in young children and road traffic accidents in older children) (4). Shewmon (10) also suggests that children may continue to recover long after adults have reached a plateau, possibly due to their retained potential for further growth and development. They may also be more likely to be offered long-term life support than adults because of the fact that people see them as more 'worthwhile' of the use of resources (4). Some may believe that death is preferable to survival with a severe disability, but Shewmon (10) insists that children and their families adapt better to physical and mental disabilities than most adults do. However, many of these views were difficult to substantiate as children and neonates tend to be excluded from studies on the basis that an accurate and consistent diagnosis of a VS is difficult to make, especially in the youngest children due to a limited capacity to show higher cognitive function (3,14). This is a complex issue with a small amount of applicable research literature available, making it difficult to conclude if any advantage is offered by a younger age. At the other extreme of life, superficially it would appear that there is a relationship between increasing age and mortality in the VS (15), but after adjustment for the severity of the illness and co-morbidity it no longer appears to exist (11,15). Although the age of the patient does not seem to directly contribute to the prediction of outcome from the VS, it may act as a substitute for other important and otherwise unmeasurable cofactors that do (15) such as pre-existing co-morbidity (e.g. cardiovascular disease) and reduced physiological reserve. Therefore, age should still be taken into account in prediction, but not as the deciding factor in where rationing of resources may be an issue and could lead to claims of ageism.

RECOVERY FROM THE VS

There are two dimensions of recovery from the VS (4):

1. Recovery of consciousness

Verified by consistent evidence of self and environment, interaction with others and voluntary behavioural responses.

2. Recovery of function

Characterised by communication, ability to learn and perform adaptive tasks, mobility, self-care and participation in recreational or vocational activities.

Recovery of consciousness may occur without recovery of function, but the converse is never true (4). The VS may be a transient stage in the recovery from coma or it may persist until death (6). The average duration of survival is 2-5 years and mortality for adults has been quoted at 82% at three years and 95% at five years (3,4). Studies have shown that the duration of survival is similar between children and adults (5-7 years), but in infants under one year it is much shorter, with estimates at a maximum of four years (4,8). There have been cases reported of patients being alive in the VS many years after this (48 years is the longest known case) (7), but the probability of prolonged survival in the VS (i.e. longer than fifteen years) has been estimated at less than 1 in 15,000 to 75,000 (4). The cause of death in the VS is most commonly infection (usually of the respiratory or urinary tract) or generalised systemic failure. Underlying comorbidity (such as ischaemic heart disease) and other unknown causes also claim a small proportion of patients but exact figure are not recorded (4).

OTHER ANCILLIARY TESTS

Other tests alone can neither diagnose nor predict if the VS will be permanent. However, when used in conjunction with the clinical examination, they can provide useful supportive information (3).

Imaging Studies (Neuroimaging)

There are no established patterns seen on neuroimaging that have been proven to predict outcome (3,9) and Bates (2) feels that their value in prediction is no better than that of the basic clinical signs. Neuroimaging methods often document lesions so severe and diffuse that awareness is highly improbable, given our current understanding of the anatomy and physiology of the brain. Several studies have documented that patients with serial abnormal scans do not recover consciousness and have progressive brain atrophy (3,7).

Magnetic resonance imaging is proven to be more sensitive than computerized tomography (CT) for detection of traumatic and ischaemic cerebral lesions (9). Kampfl et al. (9) found that although multiple lesions were commonly seen in the VS, additional injuries to specific parts of the brain were of particular importance to its persistence. Patients with lesions of the corpus callosum and dorsolateral upper brainstem had a 214-fold and 7-fold higher probability, respectively, for non-recovery (i.e. VS becoming permanent). However, the study sample size was too small to definitively prove these findings (9,16).

Cerebral Metabolic and Blood Flow Studies

Functional assessment of brain activity has been investigated as an aid to prediction. There is evidence that cerebral blood flow (CBF) is decreased in patients in established VS, with estimates at 10-50% of normal (3,9,17). However, it is accepted that measurement of CBF in the acute phase is of no prognostic significance (3) and it does not reliably predict if recovery is possible.

Cerebral metabolic activity has also been implicated in the prediction of outcome. Using positron emission tomography, a collection of studies (3,9) demonstrated a decreased cerebral metabolic rate of only 40-60% of normal in 20% of adults in permanent VS (3). Unfortunately, the limited power of these studies (due to small sample sizes) means that as yet, there is insufficient evidence to incorporate cerebral metabolic rates or CBF studies into routine practise (3,7,9).

Electrophysiology: EEGs and Evoked Potentials

The transition from coma to the vegetative state is not characterised by obvious EEG changes; it is a clinical diagnosis (2). Most patients in persistent VS show diffuse generalised polymorphic delta or theta activity, which is un-reactive to sensory stimuli. In other patients, alpha activity is the most obvious EEG feature, or some low background activity is all that can be detected (2,3). Epileptiform and seizure activity is rare in VS.

A similar pattern is seen in children although the EEG activity may be of a lower voltage and more discontinuous (3). Before three months, the EEG pattern is termed 'neonatal' and is very different to the EEG seen after this time ('mature' EEG) and throughout the rest of adulthood. This transition is termed encephalisation and is when the child's brain changes from mainly reflex subcortical functioning to cortically mediated cognition. Once this change has taken place, the same prognostic patterns apply to children as to adults, but before this, little information can be gained from EEG due to a lack of research on the topic (7,10). Recovery from the VS may be seen on EEG recordings as decreasing delta and theta activity and reappearance of a reactive alpha rhythm. However, this pattern has also been seen in patients who clinically remain in the VS, suggesting that it is not always predictive of recovery (3).

Sedative, anticonvulsant or anaesthetic drugs may cause depression of brain activity and lead to misdiagnosis on EEG (11). This again emphasises the need to eliminate all reversible causes of coma and to repeatedly assess the patient. However, the technical problems of performing such a measure in a busy intensive care ward with numerous potential sources of electrical interference can cause much of its practical value to be lost (2).

The most sensitive and reliable form of evoked potentials (EPs) in both adults and children are somatosensory evoked potentials (SSEPs) (3). Many studies (2,3,14,18,19) have produced consistent results as to their value but their clinical use is not yet widely adopted due to the belief that it is difficult to obtain accurate and reliable results in the intensive care setting (2,18). The advantage SSEPs have over EEGs is that sedating drugs minimally affect them (18). Many of the studies into SSEPs are not fully blinded, therefore any predictions of death or disability have been criticised as to being, to some extent, a self-fulfilling prophecy. Sleigh et al. (18) carried out a fully blinded study to counteract these claims and to show that SSEPs could be recorded in an ICU where dedicated neurophysiological personnel are not present. They found that bilaterally normal SSEPs were associated with a good outcome (i.e. recovery from the VS) but if they were of reduced amplitude, slowed conduction time or absent altogether, the prognosis was poor. Indeed, if SSEPs were absent bilaterally this carried the worse prognosis, being highly predictive of failure to regain consciousness (i.e. death or permanent VS). However, they were less predictive in traumatic aetiologies due to the structural damage that can ensue during the brain insult. Chen et al. (19) agreed with these findings, giving the poor prognostic factors of low amplitude or absent SSEPs positive predictive values of 100% and 89% respectively. However, as with their EEG findings, whereas negative factors were highly specific, positive ones were not, meaning that although absent or delayed SSEPs accurately predicted a poor outcome, normal SSEPs did not automatically predict a good recovery.

It has been suggested by Yingling et al. (20) that some brainstem EPs can be of predictive value, in particular P300 evoked responses. They suggest that the presence of P300 could indicate the integrity of brain systems that mediate cognitive functions, even in the absence of consciousness or overt behavioural responses. Unfortunately, the work that they have published to show its value was on a very small number of patients and therefore does not have adequate power to alter clinical practise at this time. Its diagnostic value is also limited by the fact that brainstem auditory EPs have been shown to be preserved when SSEPs are absent meaning that the predicted outcome does not alter from performing SSEPs alone - the best outcome possible is survival in the VS or death (3). Therefore, the presence of P300 cannot necessarily be correlated with outcome.

Responses to Stimuli

Motor or eye movements and facial responses such as grimacing in response to various stimuli are commonly seen stereotyped patterns. They are reflexive responses mediated at deep subcortical levels rather than as learned voluntary acts. Therefore, they do not indicate any degree of awareness and the family of the VS patient should be informed of the possibility of their occurrence, to prevent the creation of false hopes (3,7).

Genetic Factors

A small number of studies have suggested that some people may have a genetic predisposition to poor outcome from traumatic injury. The 4 allele of apolipoprotein E (apoE) has been shown to be associated with increased mortality. Sorbi et al. (21) stated that deposition of amyloid -protein (A) in the brain occurs in one third of individuals who die shortly after a severe brain injury. They found that the apoE- 4 allele occurred in a higher frequency in those who did not recover consciousness within a month (entered a persistent VS). For those who did recover, the frequency of the allele was comparable to that of the control population, suggesting a genetic susceptibility to the fatal effect of a head injury. However, their follow up was only for this month, so it is unclear as to whether the frequency of the allele was any higher in those who entered permanent VS. Also, as their study only included 16 patients, their findings have very little statistical power and are therefore inconclusive for the time being.

CONCLUSIONS

Predicting the outcome from the VS is a difficult challenge that, as yet, does not appear to be resolved. Factors mainly influencing prognosis remain etiology and duration. Traumatic brain insults have a better prognosis, nearly half of the patients studied recovered consciousness within six months, compared to only 15% of non-traumatic etiology (4). Patients were also more likely to make a good recovery from traumatic aetiologies in comparison to non-traumatic (4,11). Recovery after a year was rare in both groups (2,4,11). From the work considered, there is a possibility that age may come into the equation. This is largely inconclusive in the younger age groups, but in the older patients, age may act as a substitute for other unmeasurable factors such as comorbidity, and thus should be taken into account when determining prognosis (4). The evidence

is less conclusive in children in comparison to adults, but it appears that a similar pattern is seen in terms of functional recovery. There is evidence to suggest that children may have a better chance of recovering consciousness than adults due to the developmental potential of their brains, especially in non-traumatic etiology (4,10). False predictions of poor outcome should be avoided because this may lead to withdrawal of support in patients with the potential to recover.

Attempts have been made to treat the VS by various techniques, e.g. dopaminergic agonists, direct electrical stimulation of the brain, anticholinergics, GABA agonists, catecholaminergic anatgonists, and serotonergic agonists. However, none of these have been proven to be effective, despite seemingly encouraging preliminary studies (4,7).

Therefore, the evidence would suggest that routine clinical practise in assessment of VS patients should remain repeated clinical examinations to attempt to detect any changes in their awareness. Although criticised for the possible technical difficulties in recording, EEGs and SSEPs should be performed fairly regularly during this period. A CT scan on admission is necessary, as are the standard blood tests, to detect any potentially reversible causes of the coma.

Once a state of permanent VS is reached, little improvement can be expected. Discussion needs to take place with the family to determine the level of treatment that will be given for the duration of their lives; this is mostly preventative (e.g. avoiding contractures, pressure sores, etc). The decision to withdraw artificial nutrition and hydration is not an easy one and patients usually die within ten to fourteen days of acute dehydration and electrolyte imbalance (4,7). In the UK, it remains obligatory to seek a legal ruling to do so unless there is a clear advance directive (5). However, decisions should not be taken before a year has elapsed as the chances of recovery are small but they are a realistic possibility, although the best outcome available is usually severe disability (5,12,13).

Improvement on previous studies is difficult, if not impossible. More work is needed on the subject of prognosis in the VS to allow definitive guidelines to be agreed upon but the same problems of small sample sizes and a low incidence of the condition persists. This will most likely continue to be a problem until a large scale, multicentre trial is organised with the cooperation of major neurological centres throughout the world. This may help to alleviate some of the methodological issues but will be a massive project requiring funding which may not be possible as yet.

The VS is a complex condition that ultimately can only benefit from continued large scale studies into potential treatments and methods of accurately predicting outcome.

REFERENCES

1. Jennett B, Plum F. Persistent vegetative state after brain damage: a syndrome in search of a name. The Lancet 1:734-7; 1972
2. Bates D. The prognosis of medical coma. Journal of Neurology, Neurosurgery and Psychiatry 71(Suppl 1):120-3; 2001
3. The Multi-Society Task Force on PVS. Medical aspects of the persistent vegetative state. New England Journal of Medicine 330(21):1499-1508; 1994a
4. The Multi-Society Task Force on PVS. Medical aspects of the persistent vegetative state. New England Journal of Medicine 330(22):1572-9; 1994b
5. Zeman A. Persistent vegetative state. The Lancet 350(9080):795-9; 1997
6. Working Group of the Royal College of Physicians. The permanent vegetative state. Journal of the Royal College of Physicians of London 30(2):119-121; 1996
7. International Working Party Report on the vegetative state. Royal Hospital for Neuro-disability: London, 1996
8. Ashwal S, et al. The persistent vegetative state in children: Report of the Child Neurology Society Ethics Committee. Annals of Neurology 32(4):570-6; 1992
9. Kampfl A, et al. Prediction of recovery from post-traumatic vegetative state with cerebral magnetic-resonance imaging. The Lancet 351(9118):1763-7; 1998
10. Shewmon D. Coma prognosis in children: Part II: Clinical application. Journal of Clinical Neurophysiology 17(5):467-472; 2000
11. Levy D, et al. Prognosis in nontraumatic coma. Annals of Internal Medicine 94(3):293-301; 1981
12. Andrews K. Recovery of patients after four months or more in the persistent vegetative state. British Medical Journal 306(6892):1597-600; 1993
13. Dubroja I, et al. Outcome of post-traumatic unawareness persisting for more than a month. Journal of Neurology, Neurosurgery and Psychiatry 58(4):465-6; 1995
14. Zanderbergen E, et al. Systematic review of early prediction of poor outcome in anoxic-ischaemic coma. The Lancet 352(9143):1808-12; 1998
15. Hamel M, et al. Identification of comatose patients at high risk for death or severe disability. Journal of the American Medical Association 273(23):1842-8; 1995
16. Wardlow J, Easton V, Statham P. Which CT features help predict outcome after head injury? Journal of Neurology, Neurosurgery and Psychiatry 72:188-192; 2002
17. Oder W, et al. HM-PAO-SPECT in persistent vegetative state after head injury: prognostic indicator of the likelihood of recovery? Intensive Care Medicine 17(3):149-53; 1991
18. Sleigh J, et al. Somatosensory evoked potentials in severe traumatic brain injury: a blinded study. Journal of Neurosurgery 91(4):577-80; 1999
19. Chen R, Bolton C, Young B. Prediction of outcome in patients with anoxic coma: a clinical and electrophysiologic study. Critical Care Medicine 24(4):672-8; 1996
20. Yingling C, Hosobuchi Y, Harrington M. P300 as a predictor of recovery from coma. The Lancet 336(8719):873; 1990
21. Sorbi S, et al. ApoE as a prognostic factor for post-traumatic coma. Nature Medicine 1(9):852; 1995

Rachel Elizabeth Davison is in her final year of a medical degree (MB BS) at the University of Newcastle upon Tyne in the United Kingdom

CROSSROADS: WHERE MEDICINE AND THE HUMANITIES MEET

The Evolution, Appreciation and Representation of Music

Mamatha Bhat*, B.Sc., Sharmila Udupa

INTRODUCTION

The playing and singing of music has a profound effect on everyone, whatever the style of music to which one listens. What is miraculous is the complexity of music production by the performer, and how it can reach across to the listener and evoke certain emotions. Music can bring images to mind, trigger reminiscences of times past, and bring us to our feet. Music is a blend of melody, rhythm and harmony, and rendering it requires a great deal of concentration and memory power. It is through immense coordination that these intricate patterns can be produced; the underlying neural mechanisms are most definitely complex. These mechanisms bring about the question, "how is a multifaceted task such as music represented in the brain, and how does this representation stimulate the actions necessary for music production?"

The arrival of novel neuroimaging techniques such as fMRI (functional magnetic resonance imaging) and PET (positron emission tomography) has enabled us to visualize which areas of the brain are stimulated under certain conditions. Researchers are using these techniques to satisfy their curiosity about brain functioning with respect to specific functions such as tasting, smelling, reading, and listening. The etiology of certain psychiatric or neurological conditions is also better understood using these methods, as they offer a view of brain stimulation in real time. It is in this context that researchers have been investigating musical representation in the brain.

The therapeutic benefits of music have only recently been realized. Music therapy is an aspect of alternative mind-body medicine that is gradually gaining acceptance among medical professionals (2). It is known to have a soothing psychosomatic effect on terminally ill patients (3) and reduce anxiety associated with the diagnosis and treatment of cancer (4,5). Improvements in the pain of cancer in patients following music therapy have been noted (6), and research into this effect has implied the music may be beneficial in lowering anxiety levels associated with cancer and its treatment (7). Thus, these studies suggest music therapy to be an effective holistic practice for palliative care. In one study, quality of life was higher for those subjects receiving music therapy, and their quality of life increased over time as they received more music therapy sessions; however, life expectancy did not differ between the group receiving music therapy and those without intervention (8).

Music has also been tried as an "analgesic" during procedures such as bronchoscopy (9) and colonoscopy (10), as it has a distracting effect. Wang et al. showed a decrease in preoperative anxiety and acute postoperative pain with the use of music therapy (11). Music therapy as a clinical intervention has been demonstrated to improve mood states with a variety of populations (12). But how is it that music can effect such change in one's state of mind?

Functional MRIs and PET scanning have afforded researchers a view of the brain regions that light up in response to musical stimulation. The processing of sounds in general has been known as a function of the Heschl's gyrus, a component of the temporal lobe. Since music involves a variety of aspects, such as melody, rhythm, harmony and fine motor coordination, its representation is indisputably more complex. This review article attempts to explore the various studies that have been accomplished in

* To whom correspondence should be addressed:
Mamatha Bhat
3655 Promenade Sir William Osler, Montreal, QC, H3G 1Y6.
E-mail: mbhat1@po-box.mcgill.ca

recent years with respect to the brain's representation of these various facets of music. In addition, it looks at the purpose of such musical representation in the brain within the context of human evolution.

MUSIC APPRECIATION AS AN INNATE QUALITY

Due to their similarities, there has been widespread speculation that music and language are processed in much the same manner (13). For example, both are systematic, rule-based, and have an infinite range of possibilities. Notably, both are audibly (and in the case of language, visually) discernable at varying degrees of transformation such as changes in tempo, pitch, and speaker/singer. Also, some studies support the claim that language and music perception are innate qualities in human beings (14,15,16), with one study even relating linguistic ability to genetics (17). A prelinguistic child's capacity for musical perception is equal to that of an experienced listener. Infants recognize the familiarity of melodies across pitch and tempo changes but are not as able to detect melodies that violate the rules of musical organization (13). Also, infants below the age of six months pay more attention to sequences of consonant intervals than to those of dissonant intervals, as will be further discussed. Finally, there are indications that prenatal and/or postnatal exposure to music has no effect on the sensitivity to musical discernment, and several studies even reveal infants to be fully capable of perceiving music of any genre (13). These skills correlating to mature musical perception strongly favor the idea that appreciation of music is an innate quality.

This incredible capacity for music perception at such a tender age might find its origins in the parent-infant relationship. Lullabies and children's melodies are sung to infants at a very early age, and from the moment of birth they are spoken to in singsong voices. The genre of music that the child listens to shares a number of features with children's songs from different cultures. During the neonatal period, infants prefer renditions of a song in the maternal style to a non-maternal version of the same song by the same performer. These early social influences on music reveal the intimate relationship of social relations and musical perception. Trehub (13) speculates that it is the social nature of music and its link to positive emotional states that has permitted its growth into an elaborate system.

Hauser and McDermott believe a parallel knowledge exists within music, and it is that which contributes to its intrinsic quality (18). They have primarily used animal research so as to avoid any potential prior exposure of the subject to the music. Notably, whatever response is evoked in the animals from listening to music would reflect an auditory capacity rather than an adaptive function. Rhesus monkeys, close to humans in terms of evolution, have displayed in past studies an intrinsic inclination to tonal melodies and a similar neural response as humans to dissonant chords. It is important to bear in mind that sensitivity to musical key structure is not consistent among different species, but is within species (such as primates, infants, and adults), and that behavioral expressions may not parallel neural mechanisms at all times.

EVOLUTION OF MUSIC

If music is indeed an innate quality, the next step would be to examine its evolutionary origins. Hauser and McDermott (18) offer a number of plausible suggestions on the development of music from its origins, all heavily influenced by Noam Chomsky's (19) concept of an innate "knowledge of language". This concept refers to *an unconscious set of guidelines that define grammatical judgments, comprehension, and production.*

As for its purpose, some theorists state music to have developed as a sexually selective system devised to attract mates and signal mating potential, or to more generally express emotional states (20,21). Other theorists believe music serves to guide and facilitate collaboration among groups of people (22). There exists yet another group of scientists who believe music does not have an adaptive source, and is simply an incidental consequence of unrelated perceptual and cognitive functions (23). Certain musical forms such as Indian classical music and gospel music are devotional and spiritually inclined. Among these possibilities, the one that remains prevalent in our social context is our usage of music as a means of expression. Human and nonhuman animals still manifest their emotional states via variable vocalizations, which show that music and emotional articulation have likely developed simultaneously (18).

Rhythmic discrimination is an auditory mechanism that seems to have developed over the course of evolution as well (24). Studies have shown that human infants and monkeys recognize differences in speech rhythms between two languages, thus suggesting that certain tapping domain-general auditory mechanisms may have evolved prior to human production of music as we know it (24, 25, 26).

Solid evidence of a biological basis for music comes in the form of a study by Blood and Zatorre (27). They uncovered fascinating evidence as to why music is so appealing to all. PET was used to determine the neural stimulation that forms the basis of our pleasant emotional response to music. Subjects reported chills when listening to their favorite pieces of music. This

musical euphoria was accompanied by changes in heart rate and respiration.. Cerebral blood flow increased to brain regions implicated in reward, motivation, arousal and emotion. These areas include the amygdala, the ventral striatum, the midbrain, and the structures that are activated in response to other pleasure-inducing stimuli, such as food, sex and drug abuse. This common pattern of brain circuitry, which is involved in pleasure and reward, establishes that music is linked to "biologically-relevant, survival-related stimuli"(27).

MUSICAL REPRESENTATION IN THE BRAIN

Since music processing is clearly a mental exercise, it is logical to assume that there exists a series of neural circuits involved in its perception. What is more tantalizing, however, is to think of an entire modality for its purpose. Peretz and Coltheart (28) use Fodor's (29) definition of a modality as an entity with rapid operation, automaticity, domain-specificity, informational encapsulation, neural specificity and innateness. Although these properties are more typical than necessary, information processing and domain-specificity take importance in the definition of a mental module. Using neurologically disabled individuals with specific and isolated music-related impairments (various forms of amusia), Peretz and Coltheart have provided a comprehensive and functional model strictly for music ciphering, applying the characteristics of modular organization. In their model, a neurological defect could either interfere with information flow or with a processing component, as witnessed in detailed examination of brain-damaged patients with selective defects in music processing. The model comprises over ten processing components: tonal encoding (the knowledge of scale tones within a central tone), interval analysis, and contour analysis (pitch direction between neighboring tones) all relate to pitch organization. The remaining include rhythm analysis (fragmentation of music into temporal units based exclusively on duration value), and meter analysis (extraction of an underlying temporal rhythm with reference to strong and weak beats), which deal with temporal organization; emotional expression analysis (recognition and experience of emotion within the music), musical lexicon (bank of musical lyric exposure), vocal plan formation (results in singing), associative memories (any related nonmusical information), and two more relating primarily to speech (Figure 1).

This model is unique and more inclusive than its predecessors primarily because of its extensive organization and because of its outlook on input and output. It defines input as any acoustic stimulus that can be attributed to a single source. This input then goes to

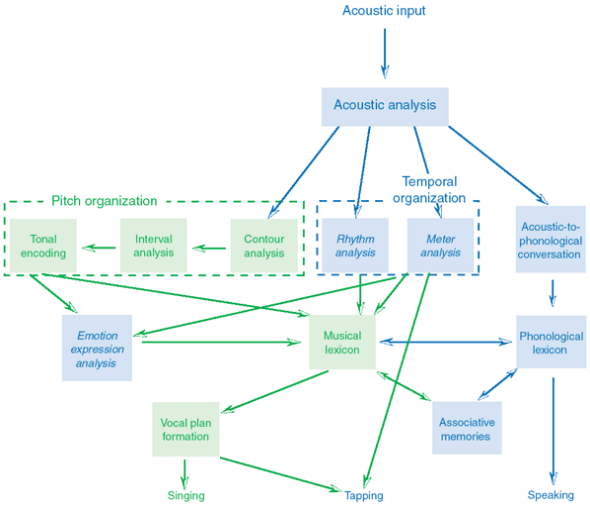


Figure 1. Modular model of music processing proposed by Peretz and Coltheart (21)

all auditory domains (not just that of music). The perceptual modules feed into an emotional analysis component, unique for each individual. Past experiences will shape what is considered emotionally appealing. The final output of the musical circuit will depend on which modality domain optimally responded to the stimulus.

A) Musical Aptitude

Perception of music has traditionally been thought of as a capacity specific to the right brain hemisphere, although a study of brain-damaged subjects by Lechevalier (30) challenged this notion. The identification and recognition of a musical piece seems to involve both hemispheres. In general, past studies have shown the left hemisphere as particularly engaged in rhythm and musical semantic representations, and the right hemisphere as specialized in melodic perception and timbre.

A study by Butcher (31) showed that auditory processing is quite different in professional musicians versus non-musicians. Primary source activity was especially localized to the anteromedial portion of Heschl's gyrus. This structure is found in the temporal lobe, and is the portion of the brain responsible for the processing of sounds. Butcher's study also detected a 130% increase in the volume of gray matter of Heschl's gyrus in professional musicians as compared to non-musicians.

A study by Hutchinson et al. discovered that the absolute cerebellar volume of musicians was significantly greater ($p < 0.001$) than that of non-musicians, a reflection of the highly developed and specialized motor skills they have developed early in

life in order to play their instruments (32). The cerebellar volume increased with reported intensity of practice, i.e. the number of hours practiced.

B) Pitch

The pitch of a note is an indication of how high or low it is in the frequency range (33). The primary auditory cortex (left temporal superior gyrus of Heschl) reveals a tonotopic distribution, the anterior and lateral portions being activated for low-pitched sounds, and more medial and posterior areas being activated for high-pitched sounds (34).

Many people with musical training have the ability to quickly identify the precise position of a note in the scale without reference to any other note; this ability is known as absolute pitch. This cognitive ability seems to develop due to the interaction of genes and environmental exposure to music during childhood in particular. Such ability is seen in musicians of all traditions, due to auditory imagery and sensorimotor response codes. According to a study by Gregerson, which points toward a strong heritable component for absolute pitch, people of Eastern Asian descent have a significantly greater incidence of absolute pitch ability than those from other backgrounds (35).

When the pitch of a song changes, typical listeners show an electrophysiological response in an area of the right frontal cortex as a reflection of their on-line memory system having been reset; however, those with absolute pitch ability show no such neural activity. Rather, those with absolute pitch show activity in the posterior dorsolateral cortex when they listen to tones (36). It is likely that this is how these people label pitches to sounds readily, since this region of the brain serves an associative function.

C) Harmony

Harmony is the sound created by simultaneously occurring pitches (33). The concept of harmony is particular to Western classical music; Eastern classical music styles tend to lay emphasis on melody. With respect to harmony, studies have shown that humans are naturally more attuned to consonance than dissonance (37). Subjectively, consonance refers to two or more sound frequencies occurring simultaneously and being pleasing to the ears of the listener. In terms of semitones (a semitone being the distance between each consecutive note on the scale), consonance tends to occur when the interval size is the octave (a difference of 12 semitones), the fifth (7 semitones), or the major third. Objectively, it has been found that consonance is translated into pairs of notes where harmonics are integer multiples of the fundamental frequency or where the fundamental notes are expressible as a ratio of small whole numbers (2:1, 3:2, 4:3, and so on).

Adult listeners have rated the minor second (a difference of 1 semitone) as being the most dissonant sound. What is striking is that, in a study by Zentner et al., infants who were exposed to consonant sounds demonstrated visual fixation at the source of sound and significantly reduced their motor (37). When the sound was dissonant, the babies were more likely to cry and turn away from the music source, thus supporting the argument for a natural human inclination towards consonance.

Harmony has been found to localize to the right lingual gyrus and the left inferior parietal lobule (38). The activation of these areas is significantly greater in musicians.

D) Melody

Melody refers to the sequence of pitches. The superior temporal gyrus has been associated with melodic processing in many neuroimaging studies (38,39,40). One study employed subjects who had undergone unilateral temporal cortectomy to relieve symptoms of epilepsy (38). This enabled determination of the roles of the various temporal lobe areas in musical processing. It was recognized that a right temporal cortectomy impaired the use of contour and interval information in melody discrimination, whereas a left temporal cortectomy adversely affected only the use of interval information. When the posterior part of the superior temporal gyrus (a portion of the auditory area) was excised, the processing of pitch and temporal variation was hindered.

In a study by Schmithorst et al. (38), unharmonized melodies bilaterally activated the superior temporal gyrus to an extent significantly greater than that associated with random tones. During this comparative exercise, musicians had significantly greater activation in the inferior parietal lobules and superior frontal gyrus bilaterally. Harmonized melodies stimulated a different activation pattern when compared to unharmonized melodies. The former generated a significantly greater activation in the right lingual gyrus and left inferior parietal lobule. A few other regions such as the right fusiform gyrus, left medial occipital gyrus, left frontal gyrus and anterior cingulate gyrus also lit up when the subjects heard harmonized melodies.

The semantic familiarity with melodies was reflected in the increased activation of the left inferior frontal and superior temporal gyri in musicians. The exact role of the parietal areas in musical processing is unknown; however auditory working memory and visuo-auditory integration are thought to be the main musical functions served. It is thought by some researchers that the supramarginal gyrus in the parietal lobe is used for

working memory during melodic processing, while visuospatial processing of harmonies is performed by the angular gyrus. The familiarity of tunes has been traced to areas of right auditory association cortex, together with right and left frontal cortices (41). It would seem that these areas are involved in imagery for familiar tunes, as evidenced by this PET scanning study. This study by Halpern and Zatorre showed retrieval from musical semantic memory as being mediated by structures in the right frontal lobe, which conflicts with results from previous studies associating left frontal areas with all semantic retrieval. The supplementary motor area (SMA) seems to be implicated specifically in image generation, and would seem to provide the link between musical perception and action.

Cases of amelodia have been reported in the past. Such isolated disorders of music perception have given researchers clues as to the localization of musical components in the brain. For instance, a highly trained musician suffered an ischemic injury of the right temporal lobe, which impaired his ability to identify the melodies of popular music pieces. He was also unable to identify instruments being played. Thus, the right temporal lobe is critical in the perception of melody. Incidentally, this region is possibly responsible for "decoding environmental sounds, discerning emotional prosody, and identifying voices" (42).

E) Rhythm

The rhythm of music is produced by the arrangement of notes and silences of varying duration. Metre is the fundamental component of rhythm, and is represented in the anterior part of the superior temporal gyrus (43). Playing music requires the ability to maintain an internal tempo. Such maintenance of tempo is facilitated when the sequences of rhythm interval durations occur in ratios of simple integers (1:2 or 1:3). These are easier to assimilate and reproduce than rhythmic sequences with more complex meters. It is thought that such simple ratio rhythms induce internal clocks or neural oscillators that assist in the perception and production of these meters. The basal ganglia and the cerebellum are believed to play a pivotal role in timekeeping mechanisms.

Sequencing behavior is key in the production of music (43). The early phase of sequence learning in music involves linking individual sequential units one at a time. As musical training progresses, the learner starts to group elements into larger combinations or "chunks". These higher-order programs become hierarchically structured into a regular pattern of sub-sequences once the subject has become familiar with a piece.

Learning a sequence that is structurally complex places cognitive demands on the brain. It is not just a matter of rote memorization. Such learning involves executive processes, chiefly error monitoring and motor program structuring. The timing aspects of both perceptual and motor tasks have been found to activate regions such as the cerebellum, supplementary motor area (SMA), premotor cortex, basal ganglia and the parietal cortex. These areas of the brain are richly interconnected and form a circuit in the perception-action cycles of music production.

CONCLUSION

Music is a sensory phenomenon that elicits perceptual and emotional responses in both the performer and the spectator. The healing powers of music have come to light, and are slowly being integrated into patient care. The ability to appreciate music has evolved over the ages. What is striking are the commonalities found between perception of speech and music with respect to rhythm and tone. In a sense, one could say that musical ability overlaps with the skills involved in language and mathematics. Numerous recent studies have tried to look at human representation of music in the brain. The neural circuits involved in music are complex, and we have only begun to scratch at the surface of its phenomenal intricacies. The various aspects of music map to different areas of the brain: the semantic familiarity of melody maps to the superior temporal gyri, harmony to the right lingual gyrus, rhythm to the cerebellum and basal ganglia. Musical training seems to result in a different brain representation of music, involving a greater number of brain regions and more complex circuits. Additional research is required to be able to pinpoint the precise cortical areas recruited by the different components of music. Modular models of music processing, such as that proposed by Peretz and Coltheart, should be investigated. It is fascinating how various music components come together, and further study should be initiated to understand how this multifaceted art form is produced in such a coordinated fashion. With respect to accomplished musicians, the question arises whether their exceptional neural circuitry is already established at an early age, prior to the molding effect of musical training. A further avenue for research would be exploration of the neural circuits that fuel creativity in music. Musical forms such as jazz and Indian classical music have improvisation within a particular framework as an integral part of their rendition. Such music most definitely would result in brain activation patterns that differ from those specific to music that is played by rote memory. Music appreciation is inherent in each one of us; it is in medicine's best interest to integrate it as part of a holistic approach to the chronic care of patients.

REFERENCES

1. Dillon WP. Neuroimaging in neurologic disorders. In: Harrison's Principles of Internal Medicine 15th edition. Toronto: McGraw-Hill Publishers; 2001: 2337
2. Fugh-Berman A. Alternative Medicine. In: Harrison's Principles of Internal Medicine 15th edition. Toronto: McGraw-Hill Publishers; 2001: 50
3. Abrams A. Music, cancer, and immunity. Clin J Oncol Nurs 2001; 5: 222-4
4. Vickers AJ, Cassileth BR. Unconventional therapies for cancer and cancer-related symptoms. The Lancet Oncology 2001; 2(4)
5. Beck SL. The therapeutic use of music for cancer-related pain. Oncol Nurs Forum 1991; 18: 1327-37.
6. Bailey LM. The effects of live music versus tape-recorded music on hospitalized cancer patients. Music Ther 1983; 3:17-28.
7. O'Kelly J. Music therapy in palliative care: current perspectives. Int J Palliat Nurs 2002; 8: 130-6
8. Hilliard RE. The effects of music therapy on the quality and length of life of people diagnosed with terminal cancer. J Music Ther 2003; 40: 113-37
9. Diette GB: Distraction therapy with nature sights and sounds reduces pain during flexible bronchoscopy: a complementary approach to routine analgesia. Chest 2003; 123(3): 941-8
10. Lee DW. Relaxation music decreases the dose of patient-controlled sedation during colonoscopy: a prospective randomized controlled trial. Gastrointest Endosc 2002; 55(1): 33-6
11. Wang SM, Kulkarni L, Dolev J, Kain ZN. Music and preoperative anxiety: a randomized, controlled study. Anesth Analg 2002; 94(6): 1489-94
12. Magee WL, Davidson JW. The effect of music therapy on mood states in neurological patients: a pilot study. J Music Ther 2002; 39(1): 20-9
13. Trehub, Sandra E. The developmental origins of musicality. Nature Neuroscience. 2003. 6: 669-673
14. Nowak, M.A., N.L. Komarova, and P. Niyogi. Computational and evolutionary aspects of language. Nature 2002: 417; 611-617.
15. Mumby, P.J., J.R.M. Chisholm, C.D. Clark, J.D. Hedley, and J. Jaubert. Language rhythms in baby hand movements. Nature 2001: 413; 35-36.
16. Patel, A.D. Syntactic processing in language and music: different cognitive operations, similar neural resources? Music Percept. 1998: 16; 27-42.
17. Lai, C.S.L., S.E. Fisher, J.A. Hurst, F. Vargha-Khadem and A.P. Monaco. A forkhead-domain gene is mutated in a severe speech and language disorder. Nature 2001: 413; 519 - 523.
18. Hauser MD, McDermott J. The evolution of the music faculty: a comparative perspective. Nature neuroscience 2003; 6: 663-68.
19. Chomsky N. Knowledge of language: its nature, origin and use. (Praeger, New York, 1986)
20. Darwin, C. The Descent of Man and Selection in Relation to Sex (John Murray, London, 1871).
21. Miller, G.F. The Mating Mind (Doubleday, New York, 2000).
22. Hagen, E.H. & Bryant, G.A. Music and dance as a coalition signaling system. Hum. Nat. (in press).
23. Pinker, S. How the Mind Works (Norton, New York, 1997).
24. deCharms, R.C., Blake, D.T. & Merzenich, M.M. Optimizing sound features for cortical neurons. Science 280, 1439-1443
25. Jusczyk, P. The Discovery of Spoken Language (MIT Press, Cambridge, 1997).
26. Mehler, J. et al. A precursor of language acquisition in young infants. Cognition 29, 143-178 (1988).
27. Blood AJ and Zatorre RJ: Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. PNAS 2001; 98 (20): 11818-11823
28. Peretz I, Coltheart M. Modularity of music processing. Nature Neuroscience. 6: 688-691, 2003.
29. Fodor J. The Modularity of Mind (MIT press, Cambridge, Massachussetts, 1983).
30. Lechevalier B, Eustache F, Rossa Y: Les troubles de la perception de la musique d'origine neurologique. Paris: Masson, 1985.
31. Butcher J. Heschl's gyrus determines musical aptitude. Lancet Neurol 1(4): 207, 2002.
32. Hutchinson S, Lee LH-L, Gaab N, Schlaug G: Cerebellar Volume in Musicians http://www.musicianbrain.com/papers/Hutchinson_Cer_CerebralC.pdf
33. Appendix F: Glossary http://www.bced.gov.bc.ca/irp/music810/apf.htm
34. Zatorre RJ, Perry DW, Beckett CA, Westbury CF, Evans AC. Functional anatomy of musical processing listeners with absolute pitch and relative pitch. Proc. Natl. Acad. Sci USA 1998; 3172-77.
35. Gregersen P, Kowalsky E, Kohn N, Marvin E. Early childhood music education and predisposition to absolute pitch: teasing apart genes and environment. Am. J. Med. Genet. 2000; 98:280-82
36. Zatorre RJ. Absolute pitch: a model for understanding the influence of genes and development on neural and cognitive function. Nature neuroscience 2003; 6:692-95
37. Zentner MR, Kagan J. Infants' perception of consonance and dissonance in music. Infant behavior & development 1998; 21:483-92
38. Schmithorst VJ, Holland SK. The effect of musical training on music processing: a functional magnetic resonance imaging study in humans. Neuroscience Letters. 348(2):65-8, 2003 Sep 11.
39. Liegeois-Chauvel C, Peretz I, Babai M, Laguitton V, Chauvel P. Contribution of different cortical areas in the temporal lobes to music processing. Brain 1998; 121:1853-67
40. Platel H, Price C, Baron J-C, Wise R, Lambert J, Frackowiak RSJ, Lechevalier B, Eustache F. The structural components of music perception, a functional anatomical study. Brain 1997; 120: 229-243
41. Halpern A, Zatorre RJ. When That Tune Runs Through Your Head: A PET Investigation of Auditory Imagery for Familiar Melodies. Cerebral Cortex, 9: 697-704, 1999
42. Sparr SA: Receptive amelodia in a trained musician Neurology 2002; 59(10): 1659
43. Janata P and Grafton ST. Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. Nat Neurosci 2003; 6:682-87.

Mamatha Bhat is a 3rd year medical student at McGill University with a B. Sc. in Microbiology and Immunology who enjoys singing and playing Indian classical music on the violin in her spare time.
Sharmila Udupa is a 1st year medical student at McGill University with an avid interest in Indian music and performing Indian classical dance.

CROSSROADS: WHERE MEDICINE AND THE HUMANITIES MEET

Legal Parameters to Medical, Ethical and Professional Responsibilities: Are Doctors Appropriately Categorised as Fiduciaries

Scott Guy, B.A., L.L.B., Barbara Ann Hocking*, B.A., L.L.B., L.L.M., PhD, and Stewart Muirhead, L.L.B. (Hons.)

INTRODUCTION

The fundamental principles of medical negligence may be constant but the nature of these incidents is perpetually evolving. Across the common law world there has been a largely judicially imposed rising standard of care expected of the medical practitioner. This paper attempts to outline the various legal requirements and the duty owed by the medical practitioner as articulated by law. It focuses on the new issues facing the medical profession and asks whether a more flexible approach, one observed most particularly in the Canadian context, and one drawing on the concept of fiduciary obligation, might create appropriate legal boundaries to deal with those issues.

Taking account of the so-called 'litigation crisis' in Australia (1), our analysis examines the doctrine of fiduciary duties, as it is presently understood, and asks whether it can be developed to provide adequate legal boundaries to the professional and ethical conduct of psychiatrists in particular. Special attention is paid to the recent Australian decision of *B -v- Marinovich* (2) , and the approach taken by the court in seeking to define a fiduciary relationship between doctor and patient.

A review of comparative jurisdictions is further undertaken to support the argument that fiduciary duties can be expanded to create new standards in the context of medical negligence. By way of contrast, the paper then compares the ethical considerations arising in the legal profession in the context of their insurance arrangements, and examines the approach taken by the courts in defining those duties. The comparison is undertaken to demonstrate the court's ability to formulate the principle governing the fiduciary obligations of professionals.

FIDUCIARY DUTIES

In order to establish negligence or "fault" on the part of a medical practitioner, a fiduciary or "special" relationship is required to be present between doctor and patient. English law (3) does not appear to recognise the existence of a fiduciary relationship between doctor and patient in the same terms, for example, as one that exists between a solicitor and client. However, Canadian law (4) does acknowledge the existence of a fiduciary relationship between doctor and patient, as indeed, Canadian law has more generously accommodated the fiduciary concept in other areas of law such as indigenous rights (5). In Australia (6), duties of a fiduciary nature may be imposed on a doctor, but they are confined and do not cover the entire doctor/patient relationship. Before examining whether a fiduciary relationship exists, the first step is to ascertain the nature of a fiduciary relationship.

What is a fiduciary?

Fiduciary relationships are referred to as relationships of trust and confidence and typically encompass the trustee/beneficiary, principle/agent, solicitor/client, employer/employee and company/director relationships.

* To whom correspondence should be addressed: Barbara Ann Hocking, School of Justice Studies, Faculty of Law, Queensland University of Technology Faculty of Law, Brisbane, Queensland, Australia, 4059; Email: b.hocking@qut.edu.au

In the leading High Court decision of *Breen v. Williams* (7), it was held that while certain elements of the doctor-patient relationship were fiduciary in nature, in essence, the relationship was contractual in character, where the medical practitioner undertakes to treat and advise the patient and to use reasonable care in doing so (8).

Defining the relationship of fidiciary

There are three principles which indicate the existence of a fiduciary relationship. The first principle is that of trust and confidence. This principle reflects the fact that there may be imbalances or inequalities of power in a relationship and therefore, as a matter of public policy, the law seeks to impose protective measures that are not ordinarily imposed (9).

A further aspect of a fiduciary relationship is that one party voluntarily undertakes to act on behalf of another party. The doctor/patient relationship satisfies this requirement because "the patient is putting his health and his life in the doctor's hands" (10). The ordinary meaning of "voluntary undertaking" means an undertaking to act in the interests of another, which would imply that one party has consented to assume the role of a fiduciary. This consent would appear to be implied in the "doctor-patient" relationship because the patient has a so-called "special vulnerability". In *Breen -v- Williams* (11) it was considered that there was little difference between a relationship where there was a "voluntary undertaking" and one in which there was a "special vulnerability" .

As well as a "voluntary undertaking" a further aspect of a fiduciary relationship is that the relationship is often one that is based on an unequal distribution of power. Because of the imbalance of power that is inherent in the doctor/patient association, this relationship can give rise to exploitation, which warrants protection in the form of an imposition of a fiduciary duty (12). It has been argued that the principle of "unequal distribution of power" cuts right to the heart of the fiduciary concept in an attempt to identify exactly what it is that makes the fiduciary special (13).

The scope of the fiduciary duty

The scope of the fiduciary duty is often determined by the nature of the fiduciary relationship, or the subject matter over which the fiduciary obligations extend (14). In Canada, where the concept has been most widely employed, fiduciary duties are not confined to the exercise of power, which can affect the legal interests of the beneficiary, but can extend to the beneficiary's "human or personal interest" (15). These interests are currently outside the protection of the law relating to the conceptualisation of the fiduciary duty in Australia.

Like the solicitor, the doctor has to provide certainty for problems that may be uncertain, and both professions must provide a high standard of professional performance. The provision of legal advice to a client is akin to the exercise by a doctor of an independent medical judgment on the patient's behalf - be it making a diagnosis, recommending treatment or implying that no treatment is necessary. Clients entrust solicitors with confidential information, and the position with the patient is no different. It is therefore difficult to see why the scope of the fiduciary duty owed by a doctor to a patient is restricted to the beneficiary's legal interests, but does not extend to personal interests. The solicitor's fiduciary obligations are based very much on the considerations outlined above. In *Re Gibson* (16), the court held that the lawyer stood in a fiduciary relationship with the client and "should exercise professional judgement solely for the benefit of the client" (17). Considerations of confidence, vulnerability and the obligation that the lawyer must act in the best interests of the client are all underpinning factors giving rise to the fiduciary duty not only to legal matters but to personal matters as well.

These principles represent the recognised basis upon which fiduciary responsibilities exist. The question remains as to whether doctors should be categorised as fiduciaries. On the one hand, it could be argued that the doctor is not an appropriate candidate for fiduciary liability because there is no inequality of power between the doctor and patient. The only power the doctor is likely to receive is from acts of consent for the purpose of diagnosis and treatment.

However, on the other hand, it might be considered that with doctors' specialised training and knowledge, such medical practitioners are therefore at a particular advantage in the relationship. Yet, while an imbalance of power may exist, a patient does have the right of veto and can withdraw consent at any time. This argument would appear to be insufficient to negate the imposition of a fiduciary duty on a doctor. Support for this contention can be gained from McLachlin J in the Canadian decision of *Norberg -v- Wynrib* (18) who, when questioning whether a fiduciary relationship existed between a doctor and patient (19), observed that the medical practitioner, Dr Wynrib, was in a position of power and could exercise that power in a way that affected the interests of the patient, Ms Norberg. This position of power was accentuated by the fact that Ms Norberg was addicted to prescription drugs. Dr Wynrib:

> *had the power to advise her, to treat her, to give her the drug or to refuse her the drug. He could unilaterally exercise that power or discretion in a way that affected her interests and her status*

as a patient rendered her vulnerable and at his mercy, particularly in the light of her addiction (20).

So, all the classic characteristics of a fiduciary relationship were held to be present.

It would appear then that a fiduciary relationship is presumed to exist between the doctor and patient. In the Australian case of *B -v- Marinovich* (21), for example, Riley J highlighted the existence of a fiduciary relationship between the psychiatrist and patient and the duty of care owed by the psychiatrist as a result of this relationship.

THE DUTY AND STANDARD OF CARE

Once a fiduciary relationship has been established between doctor and patient, the law will impose on the medical practitioner a duty to exercise reasonable care and skill in the provision of professional advice and treatment (22). This duty covers all of the ways in which a doctor is called upon to exercise his or her skill and judgment (23). The duty will extend to the examination, diagnosis and treatment of the patient and the provision of information in an appropriate case (24).

There have been two diverging approaches to evaluating whether a medical practitioner has fulfilled the appropriate duty or, more particularly, standard of care. The first approach was established in the leading case of *Bolam -v- Friern Hospital Management Committee (25).* According to the *Bolam* principle, so long as the conduct of the medical practitioner conformed to accepted medical practice then he or she could not have been considered to be negligent. The Court should not impose its own standard of care in preference to that of accepted medical opinion. As the House of Lords held in *Bolam (*26), "in short the law imposes the duty of care; but the standard of care is a matter of medical judgment."

However, later cases have tended to reject this view. The appropriate standard of care, which is to be exercised by the medical practitioner, is one that is to be decided, independently, by the Court and not be reference to medical opinion. For example, the United States Supreme Court has held that it is the responsibility of the Court to decide if the medical practitioner has fulfilled his or her duty to warn patients of the risks when undergoing medical treatment (27). A similar position has been adopted in Canada where it is the Court, and not the medical profession, that decides the extent to which the risks involved in medical treatment have been adequately disclosed (28).

The High Court has followed this approach in the leading decision of *Rogers -v- Whitaker* (29). Mason CJ held that when deciding whether a doctor has been negligent, reference needed to be made to legal principles, as well as accepted professional practice. Thus, it was the responsibility of jurors and the Court to use their own common sense when determining if a medical practitioner had been negligent, as opposed to relying on a medical "expert" (30).

This latter method enables the courts to take a more interventionist and active approach in deciding issues of medical negligence. The judge has greater freedom to scrutinise the reasoning behind an expert opinion, in much the same way as judges undertake this task in other areas of professional negligence such as in solicitors' negligence cases (31). Outside the context of medical negligence, the courts have had no difficulty with the notion that commonly adopted practices may themselves be negligent (32).

LIABILITY IN THE MEDICAL FIELD

The Marinovich case in Australia

The case of *B -v- Marinovich* (33), heard before Riley J of the Northern Territory Supreme Court, reflects the more independent position Australian courts have taken when deciding issues of medical negligence. In *Marinovich* the defendant psychiatrist was found negligent for failing to warn a patient of the addictive nature of certain tranquilliser and anti-depressant medications. The plaintiff's reliance on the medication fostered a relationship of dependence with the psychiatrist, which later led to sexual intercourse. The plaintiff also experienced serious withdrawal symptoms when she had completed the course of medication. The plaintiff claimed that on frequent occasions she had asked her psychiatrist whether there were any side effects associated with taking the prescribed drugs and whether the drugs were addictive. The defendant assured her that no side effects would be experienced and that the warnings of drug dependence, which were found on the labels of the medications, were placed there by the manufacturers to "protect themselves". The plaintiff later sued the psychiatrist for medical negligence.

The finding in B -v- Marinovich

The Court found that the psychiatrist did not inform the plaintiff of the side effects and dangers associated with the drug regime he prescribed. The plaintiff was not informed of the true nature of the drugs. She was not told of the ways in which the medication was psychologically and physiologically addictive. The patient was not given the opportunity to consider and choose a different approach. The judge considered that had the plaintiff been so advised, she would have chosen a different course of treatment. The medical evidence showed that there was an alternative and

preferable course of treatment, which was accepted by the judge. Further, the Court held that the doctor encouraged the growing dependence of the patient by adopting the regime of pharmacological treatment and by fostering an inappropriately close personal relationship with his patient. The Court was of the view that, because of the doctor/patient relationship that existed between the two parties, a duty of care was cast on the defendant to inform the plaintiff of the risks involved in taking the prescribed medication.

The law therefore recognises that a doctor has a duty to warn a patient of the risks inherent in a proposed treatment. This duty is known as the doctrine of "informed consent". The medical profession should warn the patient of all "material" risks involved in a course of treatment. A material risk, according to the Australian High Court in *Rogers -v- Whitaker*, is one that a reasonable person, in the patient's position, would attach significance to. The decision in *Marinovich* suggests that it is the responsibility of the Court to reach an independent view, as to what constitutes a "material" risk, and not rely on accepted medical opinion.

The House of Lords, however, in the *Bolam* decision, appeared to leave the determination of a legal duty to inform of a "material" risk to the judgment of doctors. Yet this decision is subject to criticism. The question of to what extent a patient should be warned before consenting cannot be answered by reference exclusively to medical practice, as the patient has a right to be informed of inherent risks. In short, the medical profession cannot be a judge in its own cause. This criticism was reflected in Lord Scarman's dissenting judgement in *Bolam.* His Lordship concluded that there was room in English law for a legal duty to warn a patient of the risks inherent in a proposed treatment: should such a duty exist, its proper legal place could be considered as an aspect of the duty of care owed by the doctor to his patient.

RECENT CASE LAW ON THE STANDARD OF CARE

However, recent cases have suggested that there are situations where the courts will have regard to medical practice when determining the relevant standard of care. For example, the courts have tended to adopt a more lenient standard of care when considering the situation of elective surgery. A New South Wales Court of Appeal decision (34) indicates that the Court will, in fact, refer to current medical practices when adjudicating on whether a medical practitioner has satisfied the relevant standard of care when undertaking elective procedures. For example, in *Tan -v- Benkovic* (35), the Court held that:

> *The medical profession is best positioned to set its own standards as to appropriate professional practices in regard to what some would regard as elective procedures paid for later…Courts should not rush into areas in which subjective professional judgements predominate… (36)*

In *Tan -v- Benkovic* the plaintiff sued the defendant because of "tightness, facial asymmetry and lines on her lips" (37) following plastic surgery. The defendant surgeon promised the plaintiff that the operation would make her look "twenty years younger" (38). In determining whether there was a breach of duty, the New South Wales Court of Appeal (39) questioned whether there was a "contumelious disregard of a doctor's duty to provide adequate care" (40). It was held that the defendant surgeon's "inducements and blandishments" did not amount to a "disregard for the doctor-plaintiff relationship" (41).

In the *Tan* case, the Court of Appeal appeared to assume a more lenient standard of care than the one adopted in *Marinovich's* case. It was held that *Rogers -v- Whitaker* did not require the surgeon to inform the plaintiff of all risks associated with the proposed operation. Further, while the plaintiff was undoubtedly "upset, vexed and depressed about the determinantal side-effects of the operation" (42) this did not amount to a breach of the doctor's standard of care. A similar approach was again adopted by the New South Wales Court of Appeal in Hunter Area *South -v- Marchlewski* (43).

On the other hand, the later decision of *Presland -v- Hunter Area Health Service* (44) has again highlighted the independent role of the Court in determining the appropriate standard of care without regard to medical practice or convention. The decision would appear to confirm the approach in *Marinovich*. In Presland, the defendant medical service discharged the plaintiff from its care since it was believed that the patient was not suffering any mental or psychiatric disorder. The plaintiff subsequent killed his brother's fiancée. The plaintiff was found to be suffering from psychosis. The New South Wales Court of Appeal held that the Hunter Area Health Service should have diagnosed the plaintiff's mental illness and reasonably foreseen that physical injury would have resulted following his discharge from the health service. The decision again emphasised the important role the Court will play in determining whether the defendant satisfied the relevant standard of care.

THE DUTY TO WARN OF RISKS INVOLVED IN TREATMENT

This raises the further concern as to how much information the doctor is required to impart to his or her

patient? It is arguable that the medical practitioner is in the best position to assess what information a patient should receive. In view of both *Rogers -v- Whitaker* and *B -v- Marinovich*, it appears that a higher, or more exacting, duty is imposed upon a medical practitioner under Australian law to inform the patient of a material risk, as compared to the approach of English law.

Criticism, however, has been raised in relation to the medical profession's ability to communicate effectively and to warn patients of the risks involved in treatment. A paper delivered at the 4th annual conference of the Australian Institute of Health, Law and Ethics in July 1999 recorded that there were a number of perceived communication limitations pervasive in the medical profession and concluded that:

> *to communicate effectively; to act promptly to protect patients from poor practice; to be open about risks and variations in performance; and to admit to the errors that they are an everyday occurrence in judgment-based clinical decision-making (45).*

With the imposition of a higher but undefined standard of care requiring medical practitioners to warn of every conceivable risk in a procedure, the question needs to be raised as to whether this will lead towards defensive medicine. Professor Jones (46) considers that this argument flies in the face of common sense and experience, which suggests that private sector defendants (solicitors, accountants, surveyors, etc) would need a sharp prod from the law of tort in order to achieve acceptable levels of competence. Jones's view has been reinforced by the recent House of Lords decision in *Arthur J S Hall -v- Simons* (47) where the Court conducted a thorough review of arguments for and against the abolition of advocates' immunity. One such argument in favour of retaining immunity was that advocates were more likely to act defensively, to the detriment of the overriding duty owed to the court in favour of their own position. That argument could not be supported as there was no evidence to suggest that advocates would act defensively contrary to their duties owed to the court and client.

CONCLUSION

The legal boundaries to the ethical standards in the legal profession are found in a fiduciary relationship, which encompasses every aspect of the lawyer-client relationship. That duty is particularly high, and understandably so, where the relationship is underpinned by one of trust and confidence. It has long been accepted that the lawyer-client relationship is a fiduciary one, principally to protect beneficiaries' fiduciary interests. Because the medical practitioner rarely has financial dealings with a patient, the courts have traditionally limited the scope of fiduciary relationship away from 'personal interests'. But with the rising standard expected of all professionals, especially so with medical practitioners, why should the fiduciary duty be limited? It is surely a relationship of trust and confidence, even within the constraints of the limited time that doctors are able to spend with patients. There may be fiduciary difficulties - for example, the situation of a white doctor and Aboriginal patient in both remote and urban parts of Australia - but this points to the need for greater awareness and enhanced training to deal with different patients.

This paper has explored the indicia giving rise to a fiduciary duty and now suggests that these concepts can be developed in the Australian common law to produce a new standard for cases such as *B -v- Marinovich*. The principle of fiduciary duties, as developed in *Marinovich's* case, also has implications for such increasingly important issues as access to medical records, and the use of a patient's genetic information. These concerns will be the focus of attention in our second, related paper

REFERENCES

1. Chief Justice Paul de Jersey "Recent Developments in Australian Negligence Law: Implications for the Insurance Industry", Insurance Council of Australia State Conference, Sheraton Hotel, Brisbane, May 23, 2003, 5. De Jersey notes the "increasingly plaintiff-friendly state in which negligence law is finding itself". His Honour points to the resultant 22 percent rise in insurance premiums in the 2001-2002 financial year: de Jersey, Insurance Council of Australia State Conference, Sheraton Hotel, Brisbane, May 23, 2003, 5. He supports the recommendations in the Federal Government's so-called "Ipp Report" which calls for a tightening of the law of negligence, in favour of defendants and away from plaintiffs.
2. (1999) NTSC 127.
3 Sidaway -v- Bethlem Royal Hospital Governors, per Lord Scarman, (1985) 1 AC 871.
4. Norberg -v- Wynrib (1992) DLR 4th 449 per McLachlin J.
5. See Dorsett, S., 'Aspassin v The Queen in Right of Canada: Re-examining the Source of the Crown's Fiduciary Obligation to Indigenous Peoples' Aboriginal Law Bulletin, 78: 7, 1996.
6. Breen -v- Williams (1995-1996) per Dawson and Toohey JJ.
7. Breen -v- Williams (1995-1996) HCA 186 CLR 71 at 89.
8 Breen -v- Williams (1995-1996) HCA 186 CLR 71 at 89.
9 See Mabo -v- Queensland (No 2)(1992) 175 CLR 1 at 203. See also the imposition of a fiduciary duty in cases of mistaken payment of money: Chase Manhattan Bank -v- Israel-British Bank (1981) 1 CH 105.
10 Sidaway -v- Board of Governors of Bethlem Royal Hospital (1985) 1 AC 871 per Lord Scarman.
11. Breen -v- Williams (1995-1996) HCA 186 CLR 71 per Gummow J.
12. This was considered by the final report of the Taskforce on Sexual Abuse of Patients (25 November 1991) p11.
13. Paul Michalik, "Doctors' Fiduciary Duties" , Journal of Law and Medicine, 6: 172; 1998.
14. Breen -v- Williams (1996) 186 CLR 71 per Brennan CJ at p82.
15. Frame -v- Smith (187) 42 DLR(4th) 81 at 104, per Wilson. J, and other non-pecuniary interests.
16. Re Gibson 369 NW 2d 466 (1985).

17. Re Gibson 369 NW 2d 466 at 476 (1985).
18. Norberg -v- Wynrib (1992) 92 DLR (4th) 449 at 499.
19. Norberg -v- Wynrib (1992) 92 DLR (4th) 449 at 499. In Canada, the law of fiduciary duties is more highly developed in the area of indigenous rights.
20. Norberg v. Wynrib (1992) 92 DLR (4th) 449 at 499.
21 (1999) NTSC 127.
22. Rogers -v- Whitaker (1992) 175 CLR per Mason CJ.
23 Sidaway -v- Governors of Bethlem Royal Hospital (1985) AC 877 at 893, per Lord Diplock.
24. Gover -v- South Australia (1985) 39 SASR 543, at 551.
25. Bolam -v- Friern Hopsital Management Committee (1957) 1 WLR 582 at 586.
26. Molam -v- Friern Hospital Management Committee (1957) 1 WLR 582 at 586.
27. Canterbury -v- Spence (18) (1972) 464 F2d 772.
28. Reibl -v- Hughes (3) (1980) 114 DLR (3d) at 13.
29. (1992) 175 CLR 479.
30. Naxakis -v- Western General Hospital (1999) HCA 22; see also Rosenberg -v- Percival (2001) HCA 18 (5 April 2001).
31. Barbara Hocking and Stewart Muirhead "Warning, Warning, Warning - All Doctors!", PN, 16: 31; 2000.
32. Edward Wong Finance Co Limited -v- Johnson, Stokes & Masters (1984) AC 296.
33. (1999) NTSC 127.
34. See Tan -v- Benkovic (2000) NSWCA 295 .
35. (2000) NSWCA 295.
36. (2000) NSWCA 295 at para. [31].
37 (2000) NSWCA 295 at para. [12].
38. (2000) NSWCA 295 at para. [6].
39. This test was also adopted in Droga v. Coluzzi (2000) NSWCA 1081; Blackwell v. AAA (1997) 1 VR 182 and Grincelis -v- House (2000) 201 CLR 321.
40. (2000) NSWCA 295 at para. [7].
41. (2000) NSWCA 295 at para. [47].
42. (2000) NSWCA 295 at para. [31].
43. (2000) 51 NSWLR 268.
44. (2003) NSWSC 754.
45. See Jones, M "Fault, Negligence and the Standard of Care: A Conceptual Enquiry into Medical Critiques of Medical Litigation", Journal of Law and Medicine, (8): 68; 1999.
46. Jones M: "Liability for Psychiatric Patients: Setting the Boundaries", PN, 16: 45; 2000.
47. (1999) 3 WLR 873 per Lord Hoffman.

Dr. Barbara Hocking BA, LLB, LLM, PhD is a senior lecturer in the Faculty of Law of Queensland University of Technology in Brisbane, Australia. **Stewart Muirhead** LLB, is Master of Laws candidate at Queensland University of Technology in Brisbane, Australia. **Scott Guy** BA, LLB, is a PhD candidate at Queensland University of Technology in Brisbane, Australia.

CROSSROADS: WHERE MEDICINE AND THE HUMANITIES MEET

Dansei Konenki: Narratives of Male Menopause in Contemporary Japan

Tomoko Sakai, B.Sc. *

ABSTRACT Previous research has focused on cross-cultural comparisons of illnesses, such as depression and senile dementia, though few have studied the actual processes by which these illness categories become separated from their roots and adopted in a different locale; in anthropological terms, their "indigenization." Through anthropological fieldwork conducted from June to September of 2003, this paper explores how dansei konenki, or male menopause, has found a niche in contemporary Japan, as well as the defining features of the country that may explain this phenomenon. Based on this research, I argue that the indigenization of dansei konenki embodies a particular socio-historical moment in Japan-namely, that of the long-running economic decline in recent years.

INTRODUCTION

Dansei konenki, a literal translation of the English term "male menopause," may be a term that is not familiar to many readers. The phrase might invoke an obscure, eccentric illness that only exists in exotic, non-western countries. On the contrary, male menopause-a paradoxical term in itself-is a strictly biomedical illness category, originating in the West (a). While it has never become well known in America, the country where it was codified, it has become a very well-known disease in Japan. Intriguingly, the term has not successfully laid its roots in its place of origin, but has somehow found its appeal in contemporary Japan.

In the summer of 2002, I conducted archival and ethnographic research on the scientific construction and lived experience of dansei konenki in Japan. During this period, I interviewed several clinicians who specialize in treating dansei konenki, as well as patients afflicted with the disease. In addition, I conducted participant observation at one clinic, which specializes in the treatment of this disorder, and in the homes and work environments of several patients-one of whom I document in this paper. Through the analyses of these data, I argue that the indigenization of dansei konenki has been fueled by and is contingent upon a particular socio-historic moment in Japan. A movement embodied by an extended economic recession, and multitudinous shifts in family values, work ethics, and gender roles that the economic decline has induced in Japanese society over the past decade.

Interestingly, this illness category has become largely divorced from its roots and taken on a distinctly different set of meanings and definitions-both in the medical community and the larger society-that reverberate within the specific context of contemporary Japan. A process frequently referred to as the localization, or indigenization of an illness category in recent anthropological literature, it is manifested in multiple layers in Japanese society: first, the western, biomedical concept of dansei konenki is grafted onto local knowledge about health and the male body, and is subsequently transformed. For example, leading Japanese physicians involved in the treatment of this condition contend that a decrease in testosterone levels leads to an imbalance in the autonomic nervous system (b), triggering chronic fatigue, shortness of breath, and

* To whom correspondence should be addressed: 1727 Waverley Street, Palo Alto, CA 94301.
E-mail address: daybyday03@yahoo.co.jp.

(a) The first scientific paper discussing this disorder was published in the Journal of American Medical Association (JAMA). In its discussion, the paper notes that "the evidence enlisted in this study strongly suggests the possibility that men also suffer from menopausal syndromes typical of women, due to endocrine disruptions in climactic years."

bad peripheral circulation; and that obesity and smoking can negatively affect the circulating levels of the hormone. In contrast to the overwhelming emphasis placed on male sexuality and its decline in the western discourse of male menopause (c), its indigenized counterpart in contemporary Japan embodies distinctly different theories of causation and physical manifestations.

The economic decline has proven to be a powerful force in reshaping important social values in the everyday lives of the Japanese-middle age men in particular-such as family values, work ethics, traditional corporate structures, and gender categories. As the illness narrative of Yoshiharu Sakaguchi, a patient diagnosed with dansei konenki will show, the drastic effects of the dragging recession on middle to old-age men-such as large-scale layoffs and major transformations in the traditional structure of companies-all signify an increasingly hostile and insecure environment for this particular population. Yoshiharu Sakaguchi is an executive of a multi-national trading company and the head of the Toyama branch. He was diagnosed with konenki while preparing for early retirement, after the company decided to close the Toyama (d) branch due to budget restraints. His narrative depicts the increasing difficulties that the creators of the current economic prosperity confront today-and accordingly, the diagnosis of konenki has added onto those distresses. For example, concurrent with the mild social stigma of being a "workaholic" who has neglected fatherly duties, which are becoming idealized as male virtues, the diagnosis also symbolizes a deprivation of masculinity for Yoshiharu Sakaguchi. Thus, dansei konenki in contemporary Japan has taken on meanings that reflect the multitude of changes and its effects that have resulted from this particular socio-economic context.

THEORETICAL RELEVANCE

Much of medical anthropological literature has focused on instances where biomedical illness categories come in contact with non-western locales. While some works have focused on biomedical illness categories that face substantial resistance for public acceptance in non-western contexts, others document the acceptance and subsequent indigenization of biomedical illnesses.

In No Aging in India: Alzheimer's, the Bad Family, and Other Modern Things (1), for example, Lawrence Cohen tackles the puzzling observation of how Alzheimer's in India is largely denied to exist. Simply put, in contrast to the U.S., where the pathologies of old age can be predominantly discussed as an unemotional and isolated medical issue, Cohen found the discourse of senility and aging in India to be intensely moral, emotional and cultural. These discrepancies, Cohen argues, may explain the apparent difficulty that Indians have in acknowledging Alzheimer's and senile dementia as a disorder that afflicts their own people, where the elderly are well-cared for and very much part of the traditional Indian joint family.

Conversely, the ethnography written by Paul Farmer looks at instances where biomedical illness categories with a western origin are taken up by non-western locales and are subsequently indigenized; similar to the case of male menopause in contemporary Japan. Paul Farmer's 1993 text, AIDS and Accusation: Haiti and the Geography of Blame documents how, as the AIDS pandemic spread in Haiti-most probably by the increased contacts between Haitian sex-workers and gay tourists in the early 1980's - local meanings, as well as theories of causation and agency were grafted onto the biomedical definition of AIDS (2). In spite of efforts to educate the public about the scientific explanation and the prevention methods of AIDS, the majority of the destitute population attributed theories of voodoo curses, Haitian black magic, and white American racism to this illness in desperate attempts to assign blame and find a cure.

Like Farmer's work, the study of dansei konenki in contemporary Japan addresses the same issues of the adoption, transformation, and subsequent indigenization of an illness category with its roots in the West. However, the uniqueness of this case study lies in the fact that male menopause has become much more rapidly and readily acknowledged as a sound, biomedical illness in Japan, in contrast to the western countries where it was initially codified. In other words, this paper expands the analytic framework of Farmer by looking at the indigenization of a biomedical illness that has yet to be commonly recognized by the larger societies in the West.

(b) An unfamiliar concept in the West, jiritsushinkei shicchoushou (literally translated as 'autonomic imbalance,' is a popular term in the Japanese medical language. The condition is characterized by a broad set of physical and mental symptoms-such as headaches, insomnia, loss of appetite, vertigo, and mild depression-which all have unidentifiable or unspecific causes. Some physicians categorize the condition into three major groups: innate, psychological, and nervous-system based autonomic imbalance. (Masahiro Inode. Jiritsushinkei Shicchoushou. Takahashi shoten: 1994

(c) Although it is a topic of great interest, an elaborate discussion on the cross-cultural differences between the symptoms of male menopause from the Western and the Japanese perspective are avoided due to the limited scope of this research paper.

(d) Toyama is a small city 90 miles west of Tokyo, facing the Sea of Japan.

ETHNOGRAPHY OF ISHINKAI UROLOGY CLINIC

The observation of clinical encounters at Ishinkai Urology Clinic exemplifies how dansei konenki, as a new illness category, is diagnosed, represented, and treated in ways which are clearly not straight imports from the western biomedical discourse of this disease, where its concept originated. For example, Dr. Hiromi Yokoyama, the director of the clinic, and leading expert on the diagnosis and treatment of dansei konenki, uses an original questionnaire as his primary diagnostic tool, which patients fill out during each visit, along with a thorough medical examination, including the measurement of circulating testosterone levels. It is divided into three sections: psychological/autonomic nervous symptoms, masculinity check, and symptoms of the urinary organs. Category 1 enlists criteria such as anxiety, irritability, fatigue, depressive mood, insomnia, and hot flashes; category 2 lists the frequency of sex and sexual desires, and category 3 asks about the frequency and uncomfortable symptoms accompanying urination. The patient ranks each diagnostic criterion from 0-3, indicating the degree of severity he is experiencing.

Additionally, in contrast to defining dansei konenki as simply a consequence of declining testosterone levels, as is customary in the United States, Japanese doctors have formulated their own diagnostic requirements for this new illness category: a conspicuous example of an illness category that is re-invented in the local environment. Dr.Yokoyama notes that the myriad of symptoms patients exhibit, such as hot flashes, headaches, chronic fatigue, and lack of sexual appetite, are caused by jiritsushinkei shicchoushou (autonomic imbalance). Additionally, they must follow a cycle of ebbs and flows: if any one symptom persists, then it is not caused by menopause.

As for testosterone levels, his theory holds that the large difference between circulating free testosterone levels before and after a man enters his middle age (e) is what characterizes male menopause. Hence, rather than setting an absolute standard of hormone levels to distinguish normal from abnormal ranges - one of the principal tenets of biomedicine - diagnostic standards are set differently for each individual. Thus, there are no strict diagnostic cutoffs for what is considered to be a physiologically normal or abnormal level of testosterone (f). Many times, Dr.Yokoyama will let the patient decide on the dosage of testosterone administration, according to his subjective discretion.

Another physician pioneering the treatment of dansei konenki in Japan has his own set of standards for diagnosing the illness. A well-known cardiovascular specialist, Dr. Fuminobu Ishikura of Osaka University Medical School defines dansei konenki as a comprehensive term incorporating erectile dysfunction, cardiovascular abnormalities, and clinical depression (3). Interestingly, these two physicians come from two distinct backgrounds of medicine-Dr.Yokoyama from urology and prostate diseases, and Dr. Ishikura from the cardiovascular system. Thus, both doctors accommodate specific knowledge about their specialties to construct an illness category that fits their own definition of male menopause (g).

Treatment methods have also undergone significant indigenization and appropriation, formulated to reverberate within local contexts. The most prominent example is Dr.Yokoyama's three main methods of treatment which he almost always combines for any patient - Testosterone replacement therapy (TRT), kanpo (h), and counseling - in contrast to the standard Western treatment which is solely confined to hormone injections. Only in cases of severe clinical depression, will he refer his patients to a psychiatrist. Dr. Yokoyama also uses supplements such as fukoidan (i), a type of dietary fiber extracted from seaweed, as part of his treatment methods. Dr. Ishikura, on the other hand, asks the first-time patient to complete a comprehensive questionnaire used in diagnosing clinical depression, and requires a session of relaxation methods, more commonly known as jiritushinkei kunren hou, or training of the autonomic nervous system, in addition to drug prescription and psychological counseling.

Other means of the local indigenization of dansei konenki are manifested in Dr.Yokoyama's numerous articles and publications. In one article entitled "Male

(e) Here, I define "middle age" as the age range from the mid 40's to mid 50's.

(f) The concept of applying a relative scale of individual hormone levels to diagnose a condition is not a common practice in Japan, and is characteristic of Dr.Yokoyama's own theory and treatment methods. Most medical institutions set a reference level, which is typically around 13-15 pg free testosterone/ml blood.

(g) Dr.Yokoyama and Dr.Ishikura are two prominent physicians involved in the treatment of male menopause. Their theories, as well as their clinic/hospital work, are completely independent from one another.

(h) Traditional Chinese herbal medicine. In her ethnography East Asian Medicine in Urban Japan (U.C.Press: Berkeley, 1980), Margaret Lock describes in detail how Western medicine and kanpo co-exist in dominant medical theories, as one popular treatment method for female menopausal symptoms is the use of these Chinese herbal medicine. The uniqueness of these remedies us the emphasis placed on individual differences - in other words, combining different herbs for different individuals, even if they exhibit similar symptoms.

(i) Although research on fukoidan is still in nascent stages, some studies have been done-including those by the National Cancer Institute, which has reported the apoptosis-inducing effects of fukoidan in cancerous cells. Dr.Yokoyama is, however, unique in his use of fukoidan to alleviate konenki symptoms.

menopause, Female Menopause" he wrote for Anatani E-ru (An Eire to You) (4), a monthly subscription magazine targeted for reader audiences in their 50's. Its pages are devoted to discussing certain kinds of food that are effective in alleviating male and/or female menopausal symptoms, as well as certain personalities and professions in which there is a high or low incidence of patients with konenki. The idea of food as medicine, an emphasis on harmony, balance, and equilibrium promoting an optimal state of being are characteristic of traditionally East Asian concepts of health. In terms of daily diet, Dr.Yokoyama encourages readers to eat "sticky/slimy" food, such as fermented soybeans, a relative of the taro root, and okra, all of which supposedly have properties that "balance hormone levels, repair prostate and other male functions." Beer is discouraged because one of its main ingredients, hop, contains estrogen-like compounds (4).

Personality-wise, men with "a strong sense of responsibility, a keen sense of competition who are punctual, impatient, and always hungry for success" will have a stronger tendency to develop menopausal symptoms compared to those who are "stable in their mental states, and maintain [their] own pace of life, unaffected by the environment." These translate into vulnerable professions - those that use the brain more than the body - such as corporate workers, company executives, and those who do a lot of deskwork. By contrast, gym teachers, military personnel, and construction site workers have a lower incidence of dansei konenki (4).

With initial research beginning in the U.S. in the 1940's (5), the possibility that men may experience menopausal symptoms, and the use of the term as diagnostic, was first introduced to Japanese medical professionals at symposia held by prestigious academic institutions, such as the General Assembly of the Japan Medical Congress and the Japanese Urology Association. However, physicians involved in the treatment of konenki today claim that it received very little attention or understanding back then. Since then, the concept has gradually gained attention and acceptance from both the medical community as well as the general public - closely following the economic downfall which has resulted in socially significant events such as massive salary and job cuts among middle-age workers. By 2003, the institutions named above were seen to give much attention to the subject, and acknowledged the need for a collective effort to further new studies and reach an agreement about a standard definition, diagnosis, and treatment for this emerging illness category. Today, several university and public medical centers, such as St.Marianna Medical University and Tokyo Women's Medical University, have founded Departments of Andrology, holding outpatient hours specifically for dansei konenki.

According to one study, a random survey of 90 men between the ages 45 and 60, using a questionnaire and interview, revealed that roughly 20-40% identified with the major symptoms associated with konenki (3). One of the pioneers in the diagnosis and treatment of dansei konenki, the Department of Urology at Kansai Medical University has also reported that it had diagnosed approximately 150 patients with the disorder-roughly 90 of whom are currently receiving testosterone replacement therapy-since it founded its outpatient clinic in January 2002 (2). Unfortunately however, large-scale, formal studies investigating the general incidence of dansei konenki, as well as the precise number of physicians or medical facilities which treat the condition in Japan today have yet to be conducted. The lack of an agreement on the specifics of the disorder, including its diagnosis and treatment by the Japanese medical community, lies at the root of the problem, and efforts to expedite the formulation of standard concepts regarding this novel illness category is much needed.

Participant-observations at Ishinkai Urology Clinic suggest that dansei konenki, as an indigenized illness category in contemporary Japan, has also taken on a socially significant function: the category serves to restore the social functionality of the patient, hence contributing to the recovery of the function of his family, his company, and the dire economic climate of the society. This stands in stark contrast to the general representation of male menopause in the U.S., where the discussion is heavily centered around issues of sexuality. Indeed, many American physicians contend that male menopause is simply another way of describing sexual/erectile dysfunction in middle-age men (3).

The rapid rise in the number of male patients who are visiting konenki outpatient clinics today, reflects the collective need for such an illness category in the particular socio-economic climate of contemporary Japan. The prolonged economic recession and decline starting from the early 1990's signified the demise of old corporate rules, such as shushin koyou, guaranteed lifetime employment, and nenko joretsu, a system closely following the ingrained ideas of East Asian filial piety, whereby one's position in the company directly reflects the years of service to the organization. Once the trademarks of Japanese corporations, shushin koyou and nenko joretsu both contributed to the illusion of a man's company seeming like his alternative family. The rapid demise of old corporate models have deprived men of their job security, which was once guaranteed, inducing notions of fear, instability, and anxiety as the

burden of supporting his family continues to fall on his shoulders (6). The driving force behind the indigenization of dansei konenki, the need to give a voice to and contribute to the restoration of the man's social functionality, becomes understandable under this particular socio-economic context.

YOSHIHARU SAKAGUCHI

The following illness narrative of a 55 year old patient, illustrates how the indigenized illness category of dansei konenki has been fueled by, and is contingent upon a particular socio-economic moment in Japan, embodied by an extended economic recession and the paradigmatic shifts in family values, work ethics, and gender roles that it has induced over the past decade. Yoshiharu Sakaguchi had just received an order from the headquarters of his company to return to Tokyo after July, when the Toyama branch was scheduled to close due to radical budget cuts.

The long-term recession has also induced major structural changes that have marked a transition from the traditional to the new, "American" economic model. During the period of rapid economic growth in the 1970's, admission to a prestigious university, through a competitive entrance exam, secured one's position on the elite track. In turn, a good university name guaranteed a job in a good company, which lead to lifetime employment, or shushin koyou (j). Under this system, one's company essentially became an alternative family, where individuals worked with a sense of security and interdependency with one another. After the downward spiraling of the economy in the early 1990's and subsequent demise of this corporate model, men were suddenly facing an age of insecurity, where lay-offs became common, and finding a new job was extraordinarily difficult for middle-age men. This especially held true for individuals who had administrative positions and lacked specific marketable skills and knowledge needed for re-employment (6).

Mr. Sakaguchi suffers from a typical set of symptoms said to accompany dansei konenki: a bad case of insomnia, night sweats and hot flashes during the day, mild depression, chronic fatigue, and prostate problems, and is currently receiving testosterone injections. Mr. Sakaguchi says he was primarily relieved when everything could be finally explained and treated, although the reaction contained inherent mixed emotions. Like many patients, the diagnosis of konenki fixes the idea of a new-found anxiety by signifying a rapid and irreversible degeneration of the masculine body to the patient.

(j) This is true mostly of men, since it was not typical for women to enter the work force after school until the 1990's.

Through the process of treating his condition, Mr. Sakaguchi and his wife both acknowledge the re-establishment of a mutual understanding and new-found appreciation for each other: an instance of a favorable role that konenki has played in his life. Thus, the reverberations of this indigenized illness in the life of Mr. Sakaguchi are pluralistic and exist in constant tension with one another-unlike many one-sided theories written on the effects of indigenized biomedical diseases. The diagnosis of dansei konenki seemed to have enhanced both the quantity and quality of communication between the Sakaguchis compared to the earlier years of their marriage, when Mr. Sakaguchi was much more reticent and stoic-minded.

Indeed, their experience points to another change, induced by the demise of the old economic paradigm, that the current konenki generation has to face and adopt. Posters of a young male pop star holding his new born son under the caption "We Don't Call Men Who Don't Participate in Child-Rearing Fathers," plastered throughout the subway stations around Tokyo, indicates the idealization of men as active participants of child rearing and household duties has as another prominent feature of the recent shifts in social values. Concurrent with this change, there seems to be a mild social stigma for men of his generation, for having been a "workaholic salary-man" who has neglected fatherly duties. Thus, men like Mr. Sakaguchi are confronting changes in the modern ideals of younger men: whereas male virtues traditionally embraced values such as stoicism, financial authority, and non-cooperation of household chores, modern young men are largely expected to be open and vocal in speaking their minds, viewing the opposite sex as equally competent players in the work force, and taking an active role in domestic activities.

CONCLUSION

Through the ethnographic account of a urology clinic and the illness narrative of one patient, I have tried to illustrate the process by which dansei konenki has become largely divorced from its Western roots, and has taken on distinct definitions that reverberate within the specific context of contemporary Japan. This indigenization, I have argued, has been fueled by and is contingent upon two key social contexts: the extended economic recession and the major shifts in traditional corporate structure, work ethics, family values, and gender roles that have resulted because of this social climate.

To be sure, male menopause is a relatively minor illness in terms of the degree to which it affects the physical and mental health of the patient and his or her family, as do AIDS or Alzheimer's disease. While

illnesses that more ostensibly threaten the livelihood of patients have long been the focus of medical anthropologists writing about biomedical indigenization, a disorder like male menopause is interesting in its own right. In future studies, it would be of great interest and contribution to the field to further explore the indigenization of biomedical illness categories in traditionally non-western locales, and the socio-cultural forces that mediate and drive this process.

REFERENCES

1. Cohen, Lawrence. No Aging in India: Alzheimer's, the Bad Family, and Other Modern Things. 10-83; 1980.
2. Farmer, Paul. AIDS and Accusation: Haiti and the Geography of Blame. 1-27; 1993.
3. Ishikura et al. 47th Convention, Japanese College of Cardiology. Sept.13, 1999.
4. Yokoyama, Hiromi. Female Menopause, Male Menopause. Anatani E-ru: 22-26; Sept. 2002.
5. Heller, C.G., and Nyers, G.B. The Male Climactic: its symptomology, diagnosis, and treatment. JAMA 126: 472; 1944.
6. Onishi. Masami. The Dynamism of Japanese Human Resources Strategies. 7-132 ; 1997.

Tomoko Sakai received her B.S. degree in Biological Sciences from Stanford University in 2003, and is currently applying to medical schools. Her research interests include medical anthropology, bioethics, and international medicine.

CROSSROADS: WHERE MEDICINE AND THE HUMANITIES MEET

Lady Lazarus Revisited: Reflections of a psychiatrist on the poetry and illness of Sylvia Plath

Laszlo Varga, M.D., Ph.D.*

Psychiatric physicians have always been attracted to study art and mental illness and Sylvia Plath's tragic history is an outstanding example for such an inquiry.

The current paper received inspiration by the news that a movie has been recently completed about Sylvia Plath, and released. Therefore it was felt that some renewed reflections about her poetry and mental illness would be a timely effort. This paper is not a biography, nor any attempt for a concise analysis. Only a few selected impressions are offered here for the interested reader with some comments about "Fate" and "Determinism" in Plath's poetry. She suffered a very serious psychotic break necessitating prolonged hospitalization, but her formal diagnosis was never revealed to the public. We know that she received 26 electroconvulsive treatments which calls attention toward schizophrenia. But her general behavior as it appears in her biographies does not reveal this illness, though in her daily life she had many small bizarre elements. Schizoid personality would be one acceptable suggestion. The difficulty is buried in our recent diagnostic classification which significantly differ from that in 1963 at the time of Plath's suicide. Bipolar Disorder with a severe psychotic break would be another possibility according to our recent nomenclature, but we wish to avaoid actual compartmentalization.

Glory and fame was not attained in her lifetime. This elevated literary status arrived to her through the bitterest human deed: a reproachful abandonment of herself to death. Shocking as it may be, her suicide served as a glorification of her poetic career, creating a certain literary fame that survives. Readers, young and even mature poets, grouped around her memorial legacy, paid astounded tribute to her, elevating Plath to a kind of priestess of contemporary poetry. Ann Sexton, Marianne Moore, and no less a literary leader than Richard Wilbur paid homage to her poetic achievements. George Steiner (1) went so far as to call her most demonic poem, Daddy, the Guernica of modern poetry. Certainly, not all these idolatrous responses are appropriate, but there can be no doubt that they signify the acceptance of Plath's new style and manner of lyric writing: her air of peremptory competence. In this air, created by her, reproach, disappointment and suicide are the dominating elements in her poetry.

Looking at the treatments of mental illness, we often see the fact that the physician is at a disadvantage to help, because the patient is not the ally, but frequently the enemy of psychiatric intervention. The poems of Sylvia Plath serve as a penetrating literary example of the mental imagery of the psychotic artist, who has opened the Pandora's Box of her inner mental world.

So, so, Herr Doktor for
So, Herr Enemy
I am your opus,
I am your valuable,
The pure gold baby

That melts to a shriek.
I turn and burn.

*To whom correspondence should be addressed:Dr. Laszlo Varga, 208 Woodlawn Rd, Keene, Texas 76059

Do not think that I underestimate your great concern.

Ash, ash-
You poke and stir.
Flesh, bone, there is nothing there-

A cake of soap,
A wedding ring,
A gold filling.

Herr God, Herr Lucifer,
Beware
Beware.

(Lady Lazarus)

It is not as if other creative artists haven't suffered in the stifling grip of psychiatric illness, but only Plath has been capable of conjuring the demonic representation of her mental imagery with such a high level of artistic richness. One shudders or rebels at such an outpouring of hallucinatory fervor, which shows agonizing intelligence at the border of the unconscious, often with bone-chilling preparation for suicide and death. At least the editors of the 1960's felt so, as one after the other rejected Plath's manuscripts, with comments that the intensity of her emotions over-powered form, that her verses were "out of control", and that she was "mining a destructive art" (3). These and similar comments reached her from publishers during the last few months of her life, during which she was actually forging her richest works.

The memory of her father is an ever returning image in her poetry. His loss is an incurable wound throughout her life. Her father figure is a central theme in Plath's poetry.

A garden of mouthings. Purple, scarlet-speckled,black
The great corollas dilate, peeling back their silks.
Their musk encroaches, circle after circle,
A well of scents almost too dense to breathe in.
Hieratical in your frock coat, maestro of the bees,
You move among the many-breasted nives,
My heart under your foot, sister of a stone.

(*The Beekeeper's Daughter)*

But her father, Otto Plath, an internationally known apiarist, struggled with diabetes and, as is common with many diabetics, he neglected his condition. His toes became gangrenous and one of his legs had to be amputated. Plath was eight years old when she lost her father, never fully recovering from this tragic event.

Daddy, I have had to kill you.
You died before I had time-
Marble-heavy, a bag full of God,
Ghastly statue with one gray toe
Big as a Frisco seal.

(Daddy)

The early childhood shock is probably the breeding nidus for her reproachful attitude throughout her life and for her suicidal motives. Her first experimentation with suicide, hinted at in her own poetry, dates back to her early teen years. Once she tried to slash her wrists in a hot bath, in reminiscence of the merciful death sentences of the classical Roman emperors. At another time, she slashed her face with a kitchen knife, leaving a visible scar on her left cheek. A strange indifference to pain is noted in her poetry, surmising an almost perverted pleasure form the immersion, as if her sufferings were the appurtenances of a mystic purification.

In spite of her destructive obsession, Plath completed all her schooling with honors, with the steadfast ambition and energy of the excellent student. She always wanted to be the first and the best and usually succeeded. Her first poems were published when she was only twelve years old. Later, she completed her college Summa Cum Laude. "Fame, Fame, Fame!" burned in her imagination in neon letters. Nothing was more important to her than looking at her name and poems in print. The flame of her internal cauldron lighted her to be up and typing by 4 a.m. in order to complete her most important poetic incantations before her suicide. Knowing her end was near, she wanted to leave a completed poetic inheritance. To do so, she worked with hallucinatory fervor.

During her college years, she was perceived by her teachers as a pleasant, agreeable student, charming and feminine, with youthful energy and liveliness. They did not suspect that significant consumption of alcohol, sexual debauchery, abounding egotism and selfishness were churning under the surface. In retrospect, everyone who knew Plath personally noted that there was incomprehensible and indefinable strangeness in her. Often show was observed as overly enthusiastic, with a voluble speech peppered with mysterious comments and more than once, she created an uncomfortable atmosphere about herself. Heinz Lehmann, a Canadian professor of psychiatry, writes about this uncomfortable, uneasy aura which emanates from the behavior of the schizoid person even during symptom-free times. Sylvia Plath was a master at creating such disturbing situations. Yet, she always operated with superior intelligence and baffled some of the intellectual leaders of their time. She devoured

books and knew thoroughly about everyone important in contemporary literature, even about third-rate writers. At a tea one afternoon in Cambridge, a little known beginning poet, Lucas Meyers, asked her to dance. He listened with amazement as Plath recited his newest poem, which had been published in an obscure journal that sold only in a few dozen copies. A friendship developed between them and they corresponded for a long time.

Plath was a third year college student as Smith, when upon her arrival home from New York; she was notified that she has been rejected to take a writing class at Harvard, even though she had submitted her prize-winning short story as proof of her talent. She lost her emotional balance, tried to commit suicide again and was hospitalized. Her medical records are of confidential of course, but we do know that her condition deteriorated and she was transferred to a closed ward. She refused all contact with the outside world, and also refused to cooperate with doctors and nurses. She would not accept visitors and announced that she hated her mother. For her birthday, she received a large bouquet of yellow roses, which she immediately threw in the trash and announced "this is only for my funeral" (6). The expenses of her treatment in an exclusive hospital were covered by Mrs. Prout, a popular American novelist at that time, who established a fellowship at Smith for talented young women. Sylvia was one of the recipients of this scholarship. Still, she described her benefactor in her novel as a snobbish lesbian matron. These few examples serve only to describe the schizoid person's sense of reality.

Plath's psychiatric hospitalization lasted almost five months, in the same private sanatorium where Robert Lowell and Ann Sexton were also treated for emotional indisposition. The latter was Plath's personal friend, who also ended her life by her own hand. While in the hospital, Eric Lindemann, Chairman of Psychiatry at Harvard Medical School, supervised her therapy. Her biographers, however, were not correct in saying that Plath was humiliated or abused, because of the electroconvulsive. Nothing else helped. She would likely have remained forever in chronic care in a closed hospital ward, in a demented mental status, enticed by her hallucinatory demons if she had not received treatment.

A gray wall now, clawed and bloody.
Is there no way out of the mind?
Steps at my back spiral into a well.
There are no trees or birds in this world,
There is only a sourness.

Let us recall from medical people that about 40% of the psychotic patients do not respond to conventional anti-psychotic drugs, and Plath belonged to this group.

Plath married the English Poet Laureat Ted Hughes after a few weeks acquaintance. After lengthy travels throughout the U.S. and Europe, she and Hughes settled in England.

After seven years of marriage, in 1962, the marriage deteriorated and ended in separation. Plath's pathological and often histrionic jealousy had an important role in this; in fact, it proved to be a self-fulfilling prophesy. Hughes became involved with another woman but he probably felt an ever-increasing discomfort about his wife's theatrical exaggerations, because at the time of the separation he announced that it was impossible to live with Plath.

While married, Plath idealized her husband. A marriage to Ted Hughes meant more than love and security for her: it meant a great deal of pride, status, inspiration and future. Her husband's alienation and unfaithfulness ripped open all her vulnerable, narcissistic wounds.

Where apple bloom ices the night
I walk in a ring,
A groove of old faults, deep and bitter.

Love cannot come here.
A black gap discloses itself.
On the opposite lip.

A small white soul is waving, a small white maggot.
My limbs, also, have left me.
Who has dismembered us?

The dark is melting. We touch like cripples.

(Event)

Plath did not know her own role in the cooling of her husband's emotional attachment. She did not know the burden of living daily with a semi-psychotic person, always ready to incite a degree of unease in her interpersonal environment. This is exactly what Plath was most capable of doing in her private life.

The plain clinical truth is that her inspirational driving forces stemmed from her inner wounds. In a polemic vengeance for her lost happiness she cried out with a quotation from Virgil: "Excoriar aliquis nostris ex ossibus ultor" (7).

It is almost banal but still significant to note that after her suicide the publishers coveted exactly those writings which they had rejected earlier. Suddenly, Plath became considered to be among the most notable

authors of modern American Literature, perhaps surpassing even Emily Dickinson. However, her abandonment of the contemporary academic style created a conflagration of arguments about the significance of her poetry.

She fantasized that like the bird Phoenix rising from its ashes, she too would rise again free after suicide. She believed that afterwards she would achieve a life without inner turmoil, leaving her demons behind. In her verses, the objects of the outside world rapidly transcend to themes of hallucinatory projections, only to present astonishing, but dazzlingly bizarre images to the reader.

Out of kitchen table or chair
As if a celestial burning took
Possession of the most obtuse object -- now and then --
Thus hallowing an interval
Otherwise inconsequent
By bestowing largess, honor
One might say love.

(Black Rook in a Rainy Weather)

She led a life of self-absorption, reproaching the world, which she perceived distortedly, through the glass of her novel The Bell Jar. Sylvia Plath was an unusually complicated lyrical poet in an environment which, in those days, scorned lyrical poetry. Plath remained self-occupied, brooding over the thwarting machinations of her inner demons, who were sending undefined, but dangerous commands, ominously approaching.

But is it necessary to scrutinize the mental status of a poet while analyzing her art? Is it possible to describe her poetry as clinical? The borderline is obviously indiscernible and the rest is almost impossible to answer. Naturally, it is not necessary to know the details of Sylvia Plath's life in order to appreciate her poetry. Literary work stands on its own, whether it is pure or inflamed, sublime or vulgar, or clinical. Nevertheless, analytic scrutiny always points to and searches for connections between a work and its creator's life. Nothing is more interesting to a reader than to learn the details of an author's life in connection with the finished work. This desire is not entirely wrong, for such an interpretation can enhance our appreciation of the poetic craft. The verse is not purely the product of the creator's conscious effort; it is also the product of her unconscious motives. These details are the most exciting psychological detective work. Plath was a confessional poet, as her biographers and critics called her. She knew very well that she and all such poets, are denuded before the public eye:

The peanut-crunching crowd
Shoves in to see
The big strip tease
Gentlemen, ladies,

These are my hands,
My knees.
I may be skin and bone,
Nevertheless, I am the same, identical woman.

A miracle!
That knocks me out
There is a charge

For the eyeing of my scars
For hearing of my heart
It really goes.

And there is a charge
A very large charge
For a word or a touch
Or bit of blood

(Lady Lazarus)

These lines were written after one of her suicide attempts.

During the last year of her life, her poetry rapidly matured. Her marmoreal poetic style appeared more chiseled and elegant. Her vituperative power bore to existence such poems as the initially "ill famed", but now highly reputed, Daddy and Lady Lazarus. However, her personality became less pleasant. She expressed her opinion without inhibition (and we don't have to explain how popular someone may be who always speaks the truth). Now and again, she came close to another collapse.

The incomprehensible psychic powers reclaimed her. In early 1963 she suffered from restless insomnia, which she tried to vanquish with ever-increasing sedatives. She knew too well how to hide her troubles from her friends, and when they finally noticed her state, it was too late. They knew nothing of her first hospitalization and illness. It had been kept secret from her English acquaintances.

There are unaccustomed terrains in Plath's poetry, her inspiration rarely came from ethereal domains of the Muses, but rather from the synclinal smelting furnace of the unconscious.

The covenants of creative work and psychotic illness arrived to a final common path: the inherent predestination to an end, the inevitability towards both, the ill and the artist are attracted to each other in a

magnetic stream. There is indeed an unexplained influence in the psychotic patient within, pulling him in one and only one possible direction. He is unable to liberate himself - like the fruitless effort in sleep paralysis - from this irresistible force. There is no other way; to surrender to the psychotic delusion is an intractable necessity.

This constraint in the ill is the will in the artist. He, too, is struggling toward the inevitable, i.e. for perfection, for which he is willing to give away his life, just to make sure that one verse or line or note or brush-stroke correctly follows the others. There is a drama in this effort, pessimistic not heroic, inherited from the classical Greeks with an unmerciful "necessity" in our Western literary tradition. Therefore, Plath's suicide is not viewed as an accident or disturbance in a depressive moment, losing the Self. We respectfully disagree with A. Alvarez's interpretation (11) that she did not really want to commit suicide, that her act was only a planned cry for help. Such a style of the neurotic housewives was disdainfully undignified for her. Plath behaved at a different level of mental functioning and we believe that she planned her suicide. She had no other choice. She condemned herself in her poetry and had to follow her inner laws. She was preparing for the end; this is exactly why she worked so feverishly to complete her work. From her manuscripts, we see that she refined and finished two or three poems in a day or two, during the last few months of her life. After she reaped her harvest, completed her poems in frightful rendition, she gave herself to her demons. In Plath's history, the demons were her reality. The conclusions waited for her, as though at the end of a railroad track, with no deviation. She followed her own irreversible moral laws:

The woman is perfected.
Her dead
Body wears the smile of accomplishment,
The illusion of a Greek necessity

Flows in the scrolls of her toga,
Her bare
Feet seem to be saying;
We have come so far, it is over.

(Edge)

REFERENCES

1. The Plath Celebration: A Partial Dissent, by Irvin Howe in Sylvia Plath: The Woman and the Work, by Ed. E. Butcher, Dodd Mead & Co. New York 1977.
2. Bitter Fame, by Ann Stevenson, Houghton Mifflin, Boston 1989.
3. Introduction in Ariel Ascending, by P. Alexander, Editor by Harper & Row, New York 1955,
4. The Interpretation of Schizophrenia, by S. Arieti, by Basic Books, New York 1955 (First Edition).
5. Bitter Fame (ibid.).
6. The Bell Jar, by Sylvia Plath (pseudonym of Victoria Lucas) by Heinemann, London , 1963.
7. Virgil: Aeneid "Let someone arise from my bones as an avenger".
8. The Journals of Sylvia Plath, by Ed. F. McCullough by Dial Press, New York 1982.
9. Ariel, Poems by Sylvia Plath by Harpers & Row, New York 1961.
10. An Intractable Metal, by H. Wendler in Ariel Ascending (ibid.).
11. Sylvia Plath, A Memoir, by A. Avarez in Ariel Ascending (ibid.).

Dr. Varga received his M.D CM degree from McGill University. He was a professor of psychiatry at Loma Linda University Medical School in California. He is currently semi-retired, practicing neurology and psychiatry in the Veterans Adminstration Hospital in Dallas, Texas.

MJM
FOCUS

SPECIAL FORUM ON
TRAUMA CARE SYSTEMS

FEATURE REVIEWS

The History of Trauma Care Systems: From Homer to Telemedicine

The Evidence Supporting a Systematic Approach to the Care of the Injured Patient:From Prevention to Rehabilitation

Moishe Liberman, M.D.*†§¥, David S Mulder, M.D.†, John S Sampalis, Ph.D

FEATURE REVIEW

The History of Trauma Care Systems From Homer to Telemedicine

Moishe Liberman, M.D.*†§¥, David S Mulder, M.D.†, John S Sampalis, Ph.D.†§¥

ANCIENT SYSTEMS OF TRAUMA CARE

The systematic and organized care of injured patients was born in times of war (1). In one of the earliest human writings, Homer in the Iliad, refers to the treatment of the injured patient during the Trojan war (5th century BC) (2). Homer reports a 77% mortality rate from injury among the 147 wounded soldiers. Surgical care of these injured soldiers was poor compared to the advanced techniques of today. However, the ancient Greeks recognized the importance of systems of trauma care. Injured soldiers were transported to, and treated in, specialized barracks called klisiai or transported to offshore ships for treatment of their wounds.

Hippocrates believed that the care of traumatic injuries during war was the ideal school for surgeons. The earliest documentation of a rudimentary trauma system is the description of medical care for the Roman Legions in approximately 100 AD (3). The Romans had organized on-site first aid, ambulances and surgeons that were on-call 24 hours a day. The trauma care hospitals (valetudinarian) were strategically located near every important encampment and were fairly sophisticated in both design and concept (4).

NAPOLEONIC ERA

Dominique-Jean Larrey (1766-1842), a Frenchman, was probably the pioneer of systematic trauma care. When international war broke out in 1792, he became a field doctor in the Rhine army. While waiting in Strasbourg for action, he organized a military medical association. Once the fighting erupted, it did not take him long to realize that an organized system was needed in order to save more soldiers. He wrote (5):

> "I now discovered the trouble it took us to move our bandaging stations - our military hospitals. According to the rules, they were supposed to stay about five kilometres from the army. The wounded were left on the field until the battle was over, or gathered at some convenient spot to which the ambulance rushed. But the roads were so choked with wagons, and such delays arose, that most of the victims died before the ambulance arrived. This gave me the idea of building an ambulance that was adequate to help the wounded during the actual battle."

Following a battle at Limburg in which the conditions were awful and casualties high, Larrey wrote to the General with a proposal; he later wrote of this proposal:

> My suggestion was accepted and I received orders to construct a cart which I called the flying ambulance. My first plan was to transport the wounded on a horse-litter, but experience soon made me give it up. The next effort was to make a cart with good suspension, combining speed with safety and comfort."

Previously, wounded soldiers were left on the battlefield until the fighting ended for the day. Larrey's ambulance could evacuate these soldiers soon after injury. The ambulance carried a doctor, quarter-master, non-commissioned officer, twenty-four infantrymen, and a drummer-boy who carried the bandage kit. He replaced the saddles' pistol holders with courier bags

*To whom correspondence should be addressed: Moishe Liberman
McGill University Health Centre, Montreal General Hospital
Departments of Surgery and Clinical Epidemiology
1650 Cedar Avenue, L10-520 Montreal, Quebec, Canada H3G 1A4
phone: (514) 934-6116, fax: (514) 934-9913
e-mail: moishe.liberman@mail.mcgill.ca
†Montreal General Hospital - Department of Surgery, McGill University Health Center
§Montreal General Hospital - Department of Clinical Epidemiology, McGill University Health Center
¥ Hôpital Sacre Coeur - Department of Surgery, Université de Montreal, Montreal, Quebec, Canada

full of instruments and bandages. Larrey's "flying ambulance" was a big success. In April of 1973, Larrey was sent back to Paris with orders to arrange flying ambulances for the whole army. For his skill and efforts, Napoleon made him a Baron and the French Army's Surgeon General. Napoleon said of him: *"He is the most virtuous man I have ever known."*

Military hospitals were designed to concentrate the injured soldiers in one area and operate on them as soon as possible following injury. Larrey realized the importance of the time to definitive care on outcome and arranged to establish his military hospitals as close to the battlefields as possible. Larrey, not only organized to have the wounded evacuated from the battlefield and brought promptly to treatment centers, but was also a pioneer in expanding the role of the military surgeon to encompass all aspects of patient care (6). He was the first to realize the importance of the surgeon in organizing all aspects of the care of the injured patient - the first "trauma system". He worked to improve sanitation, procurement of food and supplies for the sick and wounded, training of medical personnel as well as the rapid evacuation of the wounded from the battlefield.

CIVIL WAR

The American Civil War was another important step in systematic care of the injured patient. The large number of casualties, primarily due to the advances made in firearms, forced the creation of an extensive infrastructure in order to support the surgeons on the battlefield and care for the injured (1). A major advance in the systematic approach to trauma care came after the war, when the Union published 'The Medical and Surgical History of the War of the Rebellion', in a six volume set (7). This national publication reported the epidemiology of injuries and mortalities that occurred during the rebellion. It also explained the techniques and system elements that were employed throughout the war.

During the war, hospitals were strategically located near creeks in order to provide water that was vital to the care of the injured soldiers. When numerous regimental hospitals were involved in a single battle, they banded together to form a single brigade hospital. The next level of treatment centre was the division hospital and the ultimate level was the general hospital (1). The Union soon recognized the deficiencies in their system of care. The small regimental hospitals were inadequate to care for the wounded. When the regiment displaced, these hospitals could not move with the regiment and transfer all the injured soldiers. This forced the establishment of independent hospitals that could receive the injured soldiers after the regiment relocated. These new hospitals were called "general hospitals", were permanent, and were able to accept the injured from the front line hospitals following displacement of the regiment (1).

WORLD WAR I

Mechanical advances were responsible for improvements in trauma care in World War I. These advances allowed for field ambulances to become motor driven, instead of horse-driven, as they had been in previous conflicts. Timely evacuation of wounded soldiers occurred through *"echelons of treatment facilities"* (8). Echelons, each with a greater treatment capacity, were established as a standard protocol. The first tier was the evacuation of injured soldiers from the frontlines by corpsmen and stretcher-bearers. Initial treatments of the wounded men were administered at battle aid stations near the battlefront. At these stations, the injured were administered narcotics, external hemorrhage was controlled and fractures were splinted. Seriously wounded men were then evacuated to clearing stations where surgeons performed emergency surgery, which consisted mostly of the debridement of wounds. Soldiers that survived were then transported to evacuation hospitals located at safe distances from the battlefields. Definitive care was delivered at these centers and patients convalesced with the ultimate goal of returning them to the front lines. This system of escalating echelons of trauma care became the foundation for modern day civilian trauma systems. Due to the huge numbers of casualties seen in some areas, the concept of triage was born. Injured patients were sorted based both on priority and salvageability.

WORLD WAR II

Emergency medical services in Britain were instituted under the direction of the Minister of Health for both civilian and British Forces in 1940 (9). The British government realised that there would be mass civilian casualties during the war and therefore the War Office and the Minister of Health agreed to pool resources in order to create a system of trauma care that made no distinction between military and civilian casualties. At the outbreak of war it was estimated that approximately 300,000 hospital beds would be needed to treat casualties. Therefore civilian hospitals, civilian physicians and allied health professionals were selected and enrolled into the British Emergency Medical Service. Furthermore, there were specific detailed guidelines established for the organization of trauma centres, their location, corridors for pre-hospital transport and triage, as well as mobile surgical teams which could be deployed close to the areas of casualties. Trauma centres were classified based on resources for

the first time in history (Adapted from - Bailey H; *Surgery of Modern Warfare*, 1942, Vol. II p.917(9)):

> Class 1A - Hospitals of over 50 beds in which full surgical facilities are available.
> Class 1B - Smaller hospitals in which there are good surgical facilities.
> Class 2 - Hospitals suitable for the treatment of convalescent surgical and chronic medical cases. In certain cases some of these hospitals were upgraded.
> Class 3 - Infectious Disease hospitals, which were kept available for their peace-time use.
> Special Hospitals - Many well-equipped special hospitals were classified between 1 and 2. In some cases they were reserved for peace-time use (e.g.: maternity, children's and mental facilities)

In World War II, the immobile medical units that were used in WWI by the United States could not keep up with the fast pace of troop movement. This need gave birth to the "*AUX units*" which were composed of special surgical teams that travelled to the front lines in order to treat wounded soldiers. Furthermore, the advent of air travel allowed for the evacuation of wounded patients by plane during the WWII conflict, which had not been previously used in war-time situations.

The passage of patients through the echelons of care established in WWI became quicker and more efficient (8). Time lag to definitive treatment was shown to have a negative impact on survival in thoracic and abdominal wounds, as well as in extremity fractures (10,11). Trueta recognized that and wrote: "Surgical aid to casualties in the frontline is impeded by many factors and has to be adapted to varying conditions, but the main basis of success is to have the wounded patient on the operating table at the earliest possible moment" (11). In WWI, the time from injury to definitive care ranged between 12 and 18 hours. This was decreased by 50% in WWII (1,12). The improvements in time to definitive care as well as the advances in antisepsis, shock resuscitation, transfusion and surgical technique contributed to significantly improved survival rates for injured patients. The many civilian physicians, surgeons and anesthesiologists who were drafted into service in WWII observed the benefits of the systematic approach to trauma care and brought back high expectations to their civilian communities in North America (8).

KOREAN WAR

The AUX units of WWII were the basis for the establishment of the MASH (Mobile Army Surgical Hospital) units utilized in the Korean conflict. The MASH unit was a mobile surgical hospital comprising 60 beds that operated to the rear of the combat area, just out of range of artillery fire. Injured soldiers no longer had to endure multiple transportations before receiving definitive care. Instead they arrived at definitive care centres often within the "golden hour" of trauma care (13). The introduction of air ambulances and helicopters were also a major advance in the timely care of the wounded in Korea. The Korean War was the first time in military history that the helicopter was used extensively to evacuate casualties from the forward battlefields to supporting medical facilities (14). These transport mechanisms reduced the time from injury to definitive care to between 2 and 4 hours and further reduced mortality to only 2.4% (1,4).

VIETNAM

The Vietnam War saw the treatment of 250,000 casualties (15). In Vietnam, due to the mountainous terrain and the consequent difficulty in evacuating injured soldiers, the helicopter was utilized extensively as a part of the pre-hospital arsenal (16,17). The first helicopters used for evacuation of injured soldiers had

5th Century B.C.	100 A.D.	Late 18th century	1914-18	1939-45	1950-53	1965-75	1960's
Ancient Greece Injured patients transported to specialised treatment centres.	Roman Legions Organise on-site first aid stations and ambulances.	Larrey (Napoleonic Army) Organizes "flying ambulance" and military hospitals.	WWI Echelons of treatment facilities improve care for injured soldiers.	WWII British organize levels of hospital care. US army develops specialized "AUX" units composed of specialized teams which traveled to front lines.	Korean War MASH units enabled injured soldiers to reach definitive care within the "golden hour".	Vietnam War Helicopter transport used rampantly to evacuate injured soldiers from front lines.	Excess civilian mortality following trauma secondary to inappropriate care observed in the US.

two pods on the outside of the aircraft on either side for evacuation of injured soldiers from the front lines to the awaiting MASH units. The classic pattern of casualty evacuation from previous conflicts was revised in Vietnam. The battalion and regimental aid stations, which had formerly been the first line of surgical care by a physician, were being systematically overflown by the medical evacuation helicopters in Vietnam that were landing in an area where definitive care could be rendered. This area was either a unit from a medical battalion, a mobile surgical hospital, a field hospital, an evacuation hospital, or a hospital ship waiting offshore. These helicopters further decreased time to definitive surgical care to between one and one and a half hours (16).

Pre-hospital time for patients treated at the U.S. Navy Hospital in Da Nang was reported to be only 80 minutes (18). In WWII, it often took four to six months from the time of injury to get an injured soldier back to the United States by hospital ship. Due to improvements in transportation as well as the newly orchestrated evacuation and treatment system, soldiers injured on the battlefields in Vietnam would often arrive at the Naval Hospital, Great Lakes, Illinois within 72 to 96 hours from the time of injury (19). The significant advances in both the systematic care of the injured patient, as well as the improvements in surgical, transfusion-related, and antimicrobial technology resulted in decreases in mortality for patients reaching medical facilities from 8% in WWI to 4.5% in WWII to 2.5% in Korea and to less than 2% in Vietnam (20,21,22). Average times to definitive care were: 10 hours in WWII, 5 hours in Korea and 1 hour in Vietnam (21).

CIVILIAN TRAUMA CARE SYSTEMS

The civilian interest and the move towards the regionalisation of trauma care in the United States were secondary to the U.S. military experience with organized trauma care (23). The care of the injured patient evolved and improved significantly in World War II and was further developed during the subsequent Korean and Vietnam wars. It was the Korean and Vietnam conflicts that provided the basis for civilian regionalised emergency medical and trauma systems (21). Civilian trauma providers learned about well-trained paramedical personnel providing care in the field, effective pre-hospital, in-hospital and pre- to in-hospital communications, rapid emergency evacuation and transport systems (helicopter evacuations), and specialized "trauma surgeons" working out of specially designed "trauma centers" or MASH units.

In the early 1960s more Americans were killed annually on the nation's highways than were killed during the entire Vietnam conflict (24). In the United States, until the late nineteen-sixties and early nineteen-seventies, trauma care mostly occurred in the city and county hospitals or at the hospital nearest to the scene of the accident (25). The hospitals receiving trauma patients were ill-equipped and ill-staffed to handle injured patients and pre-hospital care consisted of poorly trained personnel with little equipment (26). During peak hours and at night these emergency rooms were often staffed with the most junior or unprepared physicians or poorly trained "moonlighters". In the ambulance, there was often only a driver with little emergency training and the patient would be transferred unattended in the back of the ambulance to the nearest hospital. Radios were rarely available in ambulances, and when present they were mainly used to monitor police transmissions in order to try and pick up accident calls and arrive early on-scene. Rockwood recalls that in some cities throughout the US, animals received better emergency care than citizens. They had radio dispatched vehicles and well-trained personnel available for emergency calls for pets. Trauma mortality was often due to late, inadequate or unrecognized

1966	1971	1972	1973	1977	1979	1980	1984
US NRC publishes "Accidental Death and Disability" calling for improved, organised trauma care. US Congress passes "National Highway Safety Act" calling for pre-hospital coordination and communication. First civilian trauma centres established in San Francisco and Chicago.	Trauma care system established in Illinois.	8% decline in highway mortality in Illinois in the 6 months following regionalisation of trauma care.	US Congress passes "Emergency Medical Services Act" enabling regional, comprehensive EMS systems.	Levels of civilian trauma care centres established.	West, Trunkey and Lim compare outcome in a regionalised (San Francisco County) compared to a non-regionalised area (Orange County) and find decreased preventable mortality in regionalized area.	ATLS course established.	Cales shows that preventable death rate in Orange County decreases from 34% to 15% following regionalisation of trauma care.

surgical emergencies (27,28,29).

In the early 1960s, a slew of studies were published demonstrating excess mortality following trauma in non-regionalized areas. In 1961, Van Wagoner studied 606 non-combat military deaths and concluded that one sixth (103 cases) of these were secondary to injuries from which recovery could normally be expected and another one sixth from injuries which received inadequate care (96 cases) (30). This was the first published report attempting to assess preventable deaths among injured patients occurring in a non-regionalized system of care. This paper opened the eyes of healthcare providers to the poor and inadequate care that injured patients were receiving and began a movement towards establishing an effective system to prevent these needless deaths.

Following the study by Van Wagoner, Frey showed that out of 159 patients dying as a result of trauma in Michigan, which lacked a regionalized trauma system, 28 received inappropriate care (31). Gertner demonstrated that one third of deaths involving abdominal trauma following motor vehicle collisions in Baltimore, a non-regionalized area, were preventable (32) and Moylan showed that quality of care in hospitals treating trauma patients in five hospitals in Wisconsin was unacceptable in 16% of seriously injured patients (33). These preventable death studies and other reports observing excess mortality in various areas throughout North America have been vital in the move toward regionalization in respective regions (34).

The realization by the US government of the toll that trauma was taking on society, particularly young society, in terms of morbidity and mortality as well as the "*ineffective nonsystems*"(20) of trauma care led the National Academy of Sciences to dub injury the "neglected disease of modern society" (35) in the sentinel report of 1966 prepared by the Committee on Shock and Trauma of the National Research Council. This report was titled: "*Accidental Death and Disability: The Neglected Disease of Modern Society*" and nicknamed the "*white paper*". Many important and revolutionary recommendations were made which shaped trauma systems as we know them today, including: pre-hospital radio communication systems, categorization of hospitals, the development of trauma registries, implementation of hospital trauma committees, calls for research into clinical areas of trauma care and in the areas of shock and resuscitation, and injury prevention strategies. Following this vital paper, many were convinced that injury was indeed a neglected disease and that it would continue to negatively impact on society if change was not brought about. By the early 1970s, many influential members of medical society believed that lessons learned on the battlefields in Korea and Vietnam in terms of triage, rapid transport of trauma patients to definitive care centers, and standardisation of pre-hospital and in-hospital care could be applied effectively to civilian trauma patients (36).

Accidental Death and Disability significantly contributed to what we today consider standard elements of trauma care. It highlighted the importance of standards of care, protocols for pre-hospital care providers, credentialing standards for EMS providers, improvements in accident prevention, emergency first aid and medical care, ambulance services, emergency medical communication, use of air evacuation by helicopter, upgrading emergency departments, improvements and expansion of intensive care units to properly deal with injured patients and specifications for the construction of ambulances. It also called for rapid definitive care of injured patients in the hospital setting and specialized physicians specifically trained and ready at all times to take care of injured patients. This recommendation later was integral in the establishment of a new specialty in medicine - Emergency Medicine. A strong case was made for the development of a system of trauma patient care, as well as a system of subsystem components essential to the success of an overall effective effort (20). The document called for the credentialing of four different levels of hospitals to treat trauma patients and suggested that outside credentialing agencies be designated to assign these categories. One of the most important and revolutionary recommendations made in the report was that hospitals and hospital staff be accountable for the outcomes of patients under their care. The creation of trauma registries and outcome analysis, including autopsy studies were therefore born.

Based on the recommendations of *Accidental Death and Disability*, United States Congress enacted the *National Highway Safety Act of 1966*. This legislation mandated the Department of Transport to: decrease motor vehicle accident deaths, conduct research into car safety devices, to coordinate pre-hospital care and establish pre-hospital communication.

In 1971, United States Congress proposed a law consisting of program guidelines and technical assistance measures in order to create a nationally coordinated and comprehensive system of regionalized emergency accessibility and care for all American citizens (23). This led to the Emergency Medical Services Act of 1973 (37). The Act enabled the federal government to designate a lead agency role to the Division of Emergency Medical Services in order to develop regional comprehensive emergency medical service (EMS) systems. It also provided financial aid to states for the coordination of EMS activities (38).

The first civilian trauma units were established in 1966 at Cook County Hospital in Chicago and at San Francisco General Hospital in California (1,39,40,41). The first regionalized trauma system was in established in Illinois in 1971 (20,23,42,43,44,45,46,47). Lowe and Baker highlighted the concept of the "team approach" to trauma care as being of paramount importance in establishing this system of treating injured patients, which encompassed access to the system through rehabilitation (39). Hospital designation, triage and transport guidelines as well as the concept of a "burn center" were put into place. For the first time, a central bed registry and a patient distribution and triage program were established. In Illinois, there was an eight percent decline in highway mortality in the first 6 months of 1972 (following regionalisation) compared to the same six month period in 1971, prior to regionalisation (47). This decrease in injury related mortality was observed in spite of an increase in highway accidents and injuries during that same period. In 1973, R. Adams Cowley expanded the existing Shock-Trauma program at the University of Maryland to encompass the entire state and established the Maryland Institute for Emergency Medical Services (MIEMS) (48,49,50).

By 1974, only 2 states (Maryland and Illinois) had established emergency medical systems with integrated organized trauma services within these systems. However, in 1974, the trauma system concept took off and slowly, many communities started to organize trauma care. There was however, little civilian outcome data demonstrating a positive effect for systematic trauma care at that time.

In 1973, Waters reported a 38% reduction in motor vehicle accident mortality following introduction of a regionalized trauma system in Jacksonville, Florida (51). This was one of the first reports demonstrating a beneficial effect on patient outcome with a systems approach to trauma care. The system included an emphasis on pre-hospital care, well trained pre-hospital crews, rapid response times and improved pre-hospital communication.

In 1976, the American College of Surgeons Committee on Trauma (52) assumed the leadership role in trauma system development with the publication of the first edition of *Optimal Hospital Resources for Care of the Seriously Injured* (53,54). For the first time in 1977, Detmer et al. defined the four categories of hospitals designated as civilian trauma centres which were to become the basis of today's level I, II, III and IV centers (55). More equipped centers subsequently were shown to have significantly less unacceptable care compared to less equipped, or lower level centers.

WEST AND TRUNKEY REVOLUTIONIZE TRAUMA CARE

The first and landmark study critically evaluating civilian regionalized care for injured patients and comparing a regionalized to a non-regionalized area was published by West, Trunkey and Lim in 1979 (56,57). This remarkable and original study was responsible for a new field of healthcare and health services research. They retrospectively studied one hundred consecutive motor vehicle fatalities in two counties (San Francisco and Orange County) in California between 1974 and 1975. The injured patients in San Francisco County were taken to a single trauma centre and the patients in Orange County were transported to the closest receiving hospital (39 hospitals receiving injured patients). They excluded patients who were transferred from other facilities where they had received care prior to treatment in the study hospitals and patients who died prior to reaching hospital. Deaths were classified as clearly preventable, potentially preventable and not preventable by an expert panel.

Patients in Orange County were significantly younger and had injuries of lower severity than patients in the San Francisco County cohort. Nevertheless, a panel of experts deemed that thirty-seven percent (11/30) of non-CNS related deaths in the Orange County cohort were judged to be clearly preventable compared to none in the San Francisco County cohort. Another 37% (11/30) of deaths in Orange County were judged to be potentially preventable, compared to only one death in San Francisco. This study was the first to begin to shed light on the importance of specialized, early definitive care of trauma patients and the magnitude of bringing injured patients directly to appropriately staffed, experienced and equipped care facilities.

Orange County was regionalized in 1980. Following the study by West, a complementary autopsy study (58,59) was performed on patients injured in motor vehicle collisions in Orange County before and after trauma care regionalisation (60,61,62). Cales retrospectively evaluated the outcomes of patients following implementation of a regionalized trauma system in Orange County by reviewing trauma deaths via an expert panel. This was the first ever before and after study of regionalized trauma care and served as a standard to which numerous subsequent studies would be compared. Fifty-eight deaths occurring prior to regionalisation were compared to 60 deaths occurring following implementation of a trauma system. Potentially preventable death rates dropped from 34% prior to regionalisation to 15% following regionalisation ($p<0.02$). Fifty-four percent of potentially preventable deaths occurred in patients

transported to non-trauma centres, compared to 4% of patients transported to trauma centres. They also found that the death rate from vehicular trauma dropped from 15.7 per 100,000 to 13.9 per 100,000 ($p < 0.03$) in the first year following regionalisation and from 15.8 per 100,000 to 12.4 per 100,000 after 2 years of regionalisation ($p < 0.02$). These remarkable and convincing results were strengthened due to the fact that the patients in the post-regionalisation cohort had higher Injury Severity Scores (ISS) and median age compared to those in the pre-regionalisation cohort. The improvement in outcomes was in part attributed to the aggressive approach to the care of the traumatized patient following regionalisation, which was suggested by an increased percentage of patients who received surgical interventions (62). Even though there has been some debate over the statistical methods (i.e. preventable death rate analysis) used to demonstrate efficacy in the early studies of trauma systems (63,64,65,66,67,68,69), these results are not only impressive, they also are responsible for the changes in trauma care occurring over the following 30 years.

The studies out of Orange County disclosed to the public, for the first time, the problem of inadequate trauma patient care due to the absence of a system. Backed by public demand, governments and healthcare authorities were forced to be accountable for trauma outcomes to the public. The scientific evaluation of trauma systems and their impact on society by West and Trunkey from the 1970s are unparalleled in terms of both their originality and impact on trauma care systems. These studies are the basis of modern systematic trauma care as we know and take for granted today.

ADVANCED TRAUMA LIFE SUPPORT

Prior to 1980, there were no standardized protocols or programs to train physicians in the appropriate care of the injured patient. In 1976, an orthopedic surgeon from Nebraska initiated the Advanced Trauma Life Support (ATLS) Course for training physicians in trauma care, after his wife and 3 children were killed when he crashed his plane (70). The care that his injured wife and children received was poor and this motivated the surgeon to create a course in order to train physicians with little chance to practice trauma treatment skills in the acute management of injured patients. This course was revised and adopted by the American college of Surgeons Committee on Trauma in 1979. It has since become an international standardized trauma training program, further contributing to the standardization of trauma care across regions.

MODERN DAY TRAUMA SYSTEMS

In 1985 and 1988, the Committee on Trauma Research of the National Research Council and the Institute of Medicine published "*Injury in America, A Continuing Public Health Problem*" (71) and "*Injury Control, A Review of the Status and Progress of the Injury Control, Program at the Centers for Disease Control*" (72). These reports were a follow-up to the white paper and looked at the progress that had been achieved since 1966 in trauma treatment and prevention and made extensive recommendations regarding the future of trauma care and trauma systems. These recommendations were based on the extensive body of scientific evidence that had surfaced since 1966 regarding trauma system effectiveness. The committee stated that trauma was a public health problem whose toll was unacceptable. They called for the nation to address the problem through research and legislation. The challenge proposed in *Injury in America* was to establish injury prevention and treatment as a recognized interdisciplinary field of scientific evaluation and ongoing research. The 1985 report was again expanded on and reassessed in 1999 in the report put out by the Institute of Medicine; "*Reducing the Burden of Injury - Advancing Prevention and Treatment"* (73). This report re-emphasized the point that had been highlighted previously in *Accidental Death and Disability, Injury in America* and *Injury Control:* investment in injury research in the United States did not balance the magnitude of the problem of injury. It further emphasized the positive impact of systems of trauma care on the outcome of injured patients and called for the development of more trauma systems throughout the country.

Trauma systems and regionalized trauma care has seen multiple changes and improvements over the years. It is the authors' opinion that future challenges for trauma systems include the identification of specific components of trauma systems and their impact on outcome, the creation of effective tailor-made and cost-effective systems created to fit individual community needs, the creation of novel methods to assess population-based outcome following trauma, as well as the extension of the excellent results demonstrated in urban areas to the rural setting. The advent of telemedicine promises to improve trauma care in these rural and often inaccessible areas, however further research in this area is required (74,75,76). Furthermore, aircraft (helicopter and fixed wing) are being used to transfer critically injured patients from rural centres to urban tertiary trauma centres, improving systematic care and outcomes for patients injured at great distances from definitive care facilities.

REFERENCES

1. Trunkey DD. History and development of trauma care in the United States. Clinical Orthopaedics and Related Research 2000;374:36-46.
2. Homer. The Iliad. Translated by Robert Fagles New York: Penguin Books Ltd., 1990.
3. Eastman AB. Blood in our streets. The status and evolution of trauma care systems. Arch Surg 1992;127:677-681.
4. Trunkey DD. Trauma care systems. Emerg Med Clin N Am 1984;2:913-922.
5. Calne Sir Roy. The Illustrated Hsitory of Surgery. Calne Sir Roy. Second ed. Barbican, London: Harolde Starke Publishers Ltd, 2000.
6. Davis JH, Pruitt JH, and Pruitt BA. History. Chapter 1. History. Mattox KL, Deliciano DV, and Moore EE. Fourth Edition ed. New York: McGraw-Hill Companies Inc., 2000.
7. Woodward JJ. The Medical and Surgical History of the War of the Rebellion. Washington, DC: Government Printing Office, 1875.
8. Mullins RJ. A historical perspective of trauma system development in the United States. J Trauma 1999;47:S8-S14.
9. Bailey H. Surgery of Modern Warfare. Chapter LXXX. Surgery of Modern Warfare. Second Edition ed. Baltimore, Maryland: The Williams and Wilkins Company, 1942.
10. Beebe GW and DeBakey ME. Battle Casualties. Springfiel, Illinois: Charles C Thomas, 1952.
11. Trueta J. Principles and Practice of War Surgery. First Edition ed. St. Louis: C.V. Mosby Company, 1943.
12. Trunkey DD. Trauma. Scientific American 1983;249:28-35.
13. Trunkey DD. Overview of trauma. Surg Clin N Am 1982;62:3-7.
14. United States Department of Defense. Emergency War Surgery - NATO Handbook. Frank B Berry (Assistant Secretary of Defense, Health and Medical Washington, D.C.: United States Government Printing Office, 1958.
15. Champion HR. Trauma centers in the USA - past and present. Br J Surg 1999;86:721-722.
16. Eiseman B. Combat casualty management in Vietnam. J Trauma 1967;7:53-63.
17. Byerly WG and Pendse PD. War surgery in a forward surgical hospital in Vietnam: A continuing report. Military Medicine 1971;136:221-226.
18. Cloutier CT, Lowery BD, and Carey LC. The effect of hemodilutional resuscitation on serum protein levels in human hemorrhagic shock. J Trauma 1969;9:514.
19. Haacker LP. Time and its effects on casualties in World War II and Vietnam. Arch Surg 1969;98:39-40.
20. Boyd DR, Pizzano WA, Romano TL, Stiegmann GV, and Nyhus LM. Regionalization of trauma patient care: the Illinois experience. Surg Annu 1975;7:25-52.
21. Heaton LD. Army medical service activities in Viet Nam. Military Medicine 1966;131:646-647.
22. Barnes A Jr. Progree report: The blood program in Vietnam. Military Medicine 1970;135:1-7.
23. Boyd DR and Cowley RA. Comprehensive regional trauma/emergency medical services (EMS) delivery systems. World J Surg 1983;7:149-157.
24. Shackford SR. The evolution of modern trauma care. Surg Clin N Am 1995;75:147-156.
25. Blaisdell FW. The pre-Medicale role of city/county hospitals in education and health care. J Trauma 1992;32:217-228.
26. Rockwood CA Jr., Mann CM, Farrington JD, Hampton OP, and Motley RE. History of emergency medical services in the United States. J Trauma 1976;16:299-308.
27. Root GT and Christensen BH. Early surgical treatment of abdominal injuries. Surg Gynec Obstet 1957;105:264-267.
28. Perry JF and McClellan RJ. Autopsy findings in 127 patients following fatal trahhic accidents. Surg Gynec Obstet 1964;119:586-590.
29. Fitts WT, Lehr HB, Bitner RL, and Spelman JW. An analysis of 950 fatal injuries. Surgery 1964;56:663-668.
30. Van Wagoner FH. Died in hospital: a three year study of deaths following trauma. J Trauma 1961;1:401-408.
31. Frey CF, Huelke DF, and Gikas PW. Resuscitation and survival in motor vehicle accidents. J Trauma 1969;9:292-310.
32. Gertner HR Jr., Baker SP, Rutherford RB, and Spitz WU. Evaluation of the management of vehicular fatalities secondary to abdominal injury. J Trauma 1972;12:425-431.
33. Moylan JA, Detmer DE, Rose J, and Schulz R. Evaluation of the quality of hospital care for major trauma. J Trauma 1976;16:517-523.
34. Bazzoli GJ. Community-based trauma system development: key barriers and facilitating factors. J Trauma 1999;47:S22-S24.
35. National Academy of Sciences. Accidental Death and Disability: The Neglected Disease of Modern Society. Washington, DC: National Academy of Sciences / National Research Council, 1966.
36. Hackey RB. The politics of trauma system development. J Trauma 1995;39:1045-1053.
37. Law of the 93rd congress emergency medical services act of 1973. Bill/Resolution Number: Public Law 93-155. Washington DC. 11-16-1973.
38. Moore EE. Trauma sysrems, trauma centers, and trauma surgeons - Opportunity in managed competition. J Trauma 1995;39:1-11.
39. Lowe RJ and Baker RJ. Organization and function of trauma care units. J Trauma 1973;13:285-290.
40. Freeark RJ. 1982 AAST presidential address: the trauma center: its hospitals, head injuries, and heroes. J Trauma 1983;23:173-178.
41. Trunkey DD and Eastman AB. Development of Trauma Systems. Chapter 1. Development of Trauma Systems. Trunkey DD and Lewis FR. Fourth Edition ed. St. Louis, Missouri: Mosby, 1999.
42. Boyd DR. A symposium on the Illinios trauma program: a systems approach to the care of the critically injured. J Trauma 1973;13:275-276.
43. Goldberg J, Levy PS, Gelfand HM, Mullner R, Iverson N, Lemeshow S, and Rothrock J. Factors affecting trauma center utilization in Illinois. Med Care 1981;19:547-565.
44. Boyd DR, Dunea MM, and Flashner BA. The Illinois plan for a statewide system of trauma centers. J Trauma 1973;13:24-31.
45. Metzmaker CO and Folse R. A trauma center in a non-metropolitan community. J Trauma 1973;13:291-295.
46. Boyd DR. A symposium on the Illinois trauma program: a systems approach to the care of the critically injured. J Trauma 1973;13:275-276.
47. Boyd DR, Mains KD, and Flashner BA. A systems approach to statewide emergency medical care. J Trauma 1973;13:276-284.
48. Cowley RA. Trauma center. A new concept for the delivery of critical care. J Med Soc NJ 1977;74:979-987.
49. West JG, Williams MJ, Trunkey DD, and Wolferth CC. Trauma systems. Current status - future challenges. JAMA 1988;259:3597-3600.
50. Cowley RA and Scanlan E. University trauma center: Operation, design and staffing. Am Surg 1979;45:85.
51. Waters JM and Wells CH. The effects of a modern emergency medical care system in reducing automoblile crash deaths. J Trauma 1973;13:645-647.
52. Hampton OP. The Committee on Trauma of the American College of Surgeons 1922-1972. Bull Am Coll Surg 1972;57:7-13.
53. Committee on Trauma - American College of Surgeons. Optimal hospital resources for care of the seriously injured patient. 61 ed. 1976.

54. Optimal hospital ressources for care of the seriously injured. Bull Am Coll Surg 1976;61:15-22.
55. Detmer DE, Moylan JA, Rose J, Schulz R, Wallace R, and Daly R. Regional categorization and quality of care in major trauma. J Trauma 1997;17:592-599.
56. West JG, Trunkey DD, and Lim RC. Systems of trauma care. Arch Surg 1979;114:455-460.
57. West JG, Trunkey DD, and Lim RC. The Classic - Systems of trauma care. A study of two counties. Clinical Orthopaedics and Related Research 1995;318:4-10.
58. West JG. An autopsy method for evaluating trauma care. J Trauma 1981;21:32-34.
59. West JG. Validation of autopsy method for evaluating trauma care. Arch Surg 1982;117:1033-1035.
60. Cales RH. Trauma mortality in Orange County: the effect of implementation of a regional trauma system. Ann Emerg Med 1984;13:1-10.
61. Cales RH, Anderson PG, and Heilig RW. Utilization of medical care in Orange County: the effect of implementation of a regional trauma system. Ann Emerg Med 1985;14:853-858.
62. West JG, Cales RH, and Gazzaniga AB. Impact of regionalization. The Orange County experience. Arch Surg 1983;118:740-744.
63. MacKenzie EJ, Steinwachs DM, Bone LR, Floccare DJ, Ramzy AI, and The preventable death study group. Inter-rater reliability of perventable death judgments. J Trauma 1992;33:292-303.
64. Salmi LR, Williams JI, and Waxweiler RJ. Measuring the impact of trauma care on survival: rates of perventable death, effectiveness, and efficacy. J Clin Epidemiol 1990;43:399-403.
65. Mendeloff JM and Cayten CG. Trauma systems and public policy. Ann Rev Publ Health 1991;12:402.
66. Wesson DE, Williams JI, Salmi LR, Spence LJ, Armstrong PF, and Filler RM. Evaluating a pediatric trauma program: effectiveness versus preventable death rate. J Trauma 1988;28:1226-1231.
67. Roy PD. The value of trauma centres: a methodologic review. CJS 1987;30:17-22.
68. Clark RB. Trauma centers and the New Evangelists. Ann Emerg Med 1981;10:545-546.
69. Clark DE. Comparing institutional trauma survival to a standard: Current limitations and suggested alternatives. J Trauma 1999;47:S92-S98.
70. American College of Surgeons. Advanced Trauma Life Support for Doctors (ATLS). Sixth Edition ed. Chicago, IL: American College of Surgeons, 1997.
71. National Research Council and the Institutes of medicine, Committee on Trauma Research. Injury in America. A Continuing Public Health Problem. Foege WH Washington, D.C: National Academy Press, 1985.
72. National Research Council and the Institutes of medicine, Committee on Trauma Research. Injury Control, A Review of the Status and Progress of the Injury Control, Program at the Centers for Disease Control. Washington, D.C.: National Academy Press, 1988.
73. Institute of Medicine - Committee on Injury Prevention and Control, Division of Health Promotion and Disease Prevention. Reducing the Burden of Injury - Advancing Prevention and Treatment. Bonnie RJ, Fulco CE, and Liverman CT Washington, D.C.: National Academy Press, 1999.
74. Aucar J, Granchi T, Liscum K, Wall M, and Mattox K. Is regionalization of trauma care using telemedicine feasible and desirable? Am J Surg 2000;180:535-539.
75. Lambrecht CJ. Telemedicine in trauma care: description of 100 trauma teleconsults. Telemed J 1997;3:265-268.
76. Rogers FB, Ricci M, Caputo M, Sartorelli K, Callas P, and et al. The use of telemedicine for real-time video consultation between trauma center and community hospital in rural setting improves early trauma care: preliminary results. J Trauma 2001;51:1037-1041.

Moishe Liberman MD, is a fourth year resident in the Division of General Surgery at McGill University. He is also a research fellow in the Department of Experimental Surgery at McGill.
David S. Mulder MD, MSc, FRCSC is the Chairman of the Division of Cardiothoracic Surgery at the McGill University Health Center, Senior Surgeon at the Montreal General Hospital, and Professor of Surgery at McGill University.
John S. Sampalis PhD, is the Head of Surgical Epidemiology at McGill University, an Associate Professor in the Departments of Surgery, Epidemiology and Biostatistics, and an Associate Professor in the Department of Social and Preventive Medicine at Université de Montréal and at Université de Laval. He has been a past recipient of the Medical Research of Canada Scientist award and is currently a senior scientist supported by the Fonds de la recherche en sante du Quebec (FRSQ).

FEATURE REVIEW

The Evidence Supporting a Systematic Approach to the Care of the Injured Patient:From Prevention to Rehabilitation[1]

Moishe Liberman MD†, David S Mulder MD†, John S Sampalis PhD*

INTRODUCTION

Trauma is the leading cause of death for individuals under 45 years of age in the Western world and remains the fourth leading cause of death for all ages combined (1,2,3,4). Approximately 0.9 million people worldwide die secondarily to injury (8% of all deaths) (5). It is also a major cause of morbidity in both the short and long-term (6). Furthermore, injury is a leading cause of disability, potential years of life lost and a major contributor to overall health care costs (7,8,9). It is estimated that injury causes 36 life-years lost per death compared to 16 life-years for cancer and 12 years for heart disease and stroke combined (10). In 1994, 8,687 people died following accidents in Canada (1). Approximately four times as many patients suffer severe disability related to accidents each year.

The cost of acute medical care for injured patients is in excess of $16 billion per annum (11). This represents the second largest source of medical expenditures in the United States. In addition to the health dollars spent on the acute care of injured patients, an additional $150 billion US are required to cover the annual cost due to death, disability, and lost wages and taxes (9). From a health-economic perspective, the cost of trauma and its consequences makes the elucidation of evidence-based practices paramount. Trauma care systems have been shown to significantly decrease medical care costs. It is estimated that by extending trauma care systems throughout the entire United States, annual medical care payments could be lowered by $3.2 billion (12). If productivity costs due to premature death are taken into account, the total savings could total $10.3 billion.

*To whom correspondence should be addressed: Moishe Liberman
McGill University Health Centre, Montreal General Hospital
Departments of Surgery and Clinical Epidemiology
1650 Cedar Avenue, L10-520 Montreal, Quebec, Canada H3G 1A4
phone: (514) 934-6116, fax: (514) 934-9913
e-mail: moishe.liberman@mail.mcgill.ca
†Montreal General Hospital - Department of Surgery, McGill University Health Center
§Montreal General Hospital - Department of Clinical Epidemiology, McGill University Health Center
¥ Hôpital Sacre Coeur - Department of Surgery, Université de Montreal, Montreal, Quebec, Canada

TRAUMA SYSTEMS

Trauma care throughout Canada and the rest of North America has seen tremendous changes over the last 30 years. The regionalisation of trauma care, which has occurred in some Canadian and American regions, has shifted the scope of trauma patient management from hospital-based care to a systems approach. A regionalised approach to trauma care (a trauma system) consists of the global care of the injured patient, from the time of injury until the end of rehabilitation (13,14). The system provides a continuum of services encompassing four elements: [1] pre-hospital care, [2] in-hospital care [3] rehabilitation, and [4] research. The ultimate goal of these systems is to get the injured patient to definitive care as soon as possible (15,16).

Trauma systems have been designed to render "optimal care" to injured patients. Eggold defines optimal care as being based on two implied premises (17): One premise is that suboptimal trauma care is possible and demonstrable and the other premise is that optimal care must result in reduced mortality and/or morbidity, "the sine qua non of medical progress". Furthermore, by pooling resources and avoiding duplication through a system of care within a region, cost effectiveness is assured (18).

The care of injured patients is a continuum from the moment of injury, until the return to daily life (19,20). Regionalised trauma care incorporates several different elements, which together make up the trauma "system".

These systems have been repeatedly shown to decrease mortality and improve the outcome of injured patients in multiple different regions throughout the Western world.

EPIDEMIOLOGY OF TRAUMA DEATHS

Trauma is a devastating disease. It contributes to approximately 140,000 deaths per year in the United States (21). Unintentional injuries account for 4.6% of deaths and 19.6% of potential years of life lost in patients younger than 65 years of age. (22) Injuries account for 61% of deaths due to trauma in the United States and nearly half of these deaths are due to motor vehicle accidents (7). Falls, occurring mostly in the octogenarian population, account for the second most prevalent portion of unintentional deaths.

Death resulting from trauma follows a trimodal distribution (19,23,24,25,26). These peaks were first alluded to in reports by Beebe and DeBakey in 1952 (27) and by Zollinger in 1955 (28) and later expanded on by Trunkey in 1983 (19). The first peak of death following injury is dubbed the "immediate deaths" and occurs within seconds of injury. It accounts for 50% of trauma-related mortality. These early deaths occur secondary to lacerations to the brain, upper spinal cord, heart, aorta and other major vessels. Virtually all of these patients die and little, if anything, can be done to save them. Cales showed that 44% of trauma deaths occurred at the scene (29). The only way to reduce deaths in the first peak of trauma mortality is through prevention strategies and programs, as well as tougher legislation on firearms and motor vehicle traffic laws (30). Injury prevention and control has been shown to have more immediate health and economic benefits than the prevention and control of chronic diseases (31).

The second peak of mortality, the "early deaths" occurs within minutes to a few hours following injury and contributes to 30% of mortality following trauma. This period has been dubbed the "golden hour" following injury (19). Deaths in this period are secondary to injuries that require urgent and emergent care. These injuries are time-critical and the sooner the patient receives definitive care for these injuries, the better the outcome. Important injuries in this category include: subdural and epidural hematomas, hemopneumothorax, liver lacerations, ruptured spleen, pelvic and long bone fractures causing significant bleeding, as well as injuries to blood vessels contributing to significant blood loss. These injuries require timely definitive care, usually through surgery to repair the source of blood loss and stop the hemorrhage or to evacuate a compressive hematoma (cerebral hemorrhage), or an interventional procedure (tube thoracostomy, pericardiocentesis, angio-embolisation…). If these procedures are not provided promptly and properly by the appropriate personnel in the appropriate setting, mortality occurs.

It is for the patients in the second period of trauma deaths that systematic trauma care attempts to make an impact. These are the time-critical patients, desperately in need of definitive and appropriate care in a timely manner. Patients receiving rapid transport to hospital will not have good outcomes if they are taken to the wrong hospital. Patients taken to the right hospital will also have poor outcomes if there is a delay in getting them there. The second peak is the focus of trauma systems and regionalised care of the injured patient.

The third peak of mortality following trauma, the "late deaths", occurs several days or week following injury. These deaths account for approximately 20% of deaths after injury. Deaths in this period are usually secondary to sepsis and multiple organ system failure. Rapid and appropriate care can reduce these injuries, however most of these deaths will occur regardless of the system of trauma care and the key to reducing them lies in research into systemic mediators of sepsis and multiple organ dysfunction. Time is less of a factor in the outcome of these patients; rather, the quality of medical care and the state of medical knowledge contribute to outcome in these patients.

Recently researchers have identified a fourth peak of trauma deaths, which requires further study. The fourth peak of deaths is that which occurs in the first year following injury (32). The age characteristics of this unique group of patients show that patients over the age of 65 have a 15-fold greater chance of dying in the year following injury.

PROCESS OF REGIONALISATION, BUILDING A "TRAUMA SYSTEM"

The basis for the regionalisation of trauma care or the development of a "trauma system" is the need to link all aspects of care in order to maximize efficiency, pool resources and improve outcomes. A comprehensive trauma system links hospitals, pre-hospital care and other emergency medical services, post hospital care facilities (rehabilitation and long-term care centres), as well as health care and public safety agencies (33). Ideal trauma systems include prevention, access, acute hospital care, rehabilitation, and research activities (34). These systems have been developed in order to direct seriously injured patients to specific facilities on local, regional, and state/province wide bases. The two main goals of regionalised trauma care are to improve the quality of care and to decrease its cost (35).

The American College of Surgeons Committee on Trauma clearly outlines the importance of emphasising the trauma system, rather than the trauma centre as being integral in improving trauma patient outcome

(34).

> "Care of the injured patient requires a system approach to ensure optimal patient care. A systematic approach is necessary within a facility; however no one trauma centre can do everything alone. Thus, a system approach is necessary within an entire community regardless of its size…If resources for optimal care of the injured patient are to be used wisely, then some concentration of resources should occur. This type of resource allocation should allow patients to move to the highest level of care available and, ideally, should also avoid excessive and inappropriate expenditure in a time of limited medical resources."

Integral to the trauma care system is the designation of definitive trauma care facilities. These facilities provide the full spectrum of trauma care to injured patients in the most efficient and effective manner. The overall goal of the system is patient care and outcome, however efficiency and proper use of resources is emerging as an important aspect of trauma systems. Every trauma system or regionalised trauma area should have a "lead hospital". The lead hospital should be the hospital with the highest level of care (highest designation) in the area.

Trauma centres serve as the hubs of these systems. Trauma centres also exist in areas without formal trauma systems in place. In these areas they are usually not designated as trauma centres, but act as "de facto" or "functional" centres (36). Tertiary trauma centres (level I centres) are responsible for receiving the most seriously injured patients directly from the field (in most cases), as well as accepting and guiding transfer from secondary and primary centres. They also serve the purpose of being leaders in trauma care and prevention programs for the region. They are also responsible for conducting trauma-related research.

West identified eight essential elements that were integral to an inclusive trauma system based on criteria from the American College of Surgeons (37). These criteria were: (a) the presence of a lead agency with legal authority to designate trauma centres; (b) the use of a formal process for trauma centre designation; (c) the use of American College of Surgeons standards for trauma centres; (d) the use of an out-of-area survey team for trauma centre designation; (e) limiting the number of designated trauma centres in a community based on assessment of population need; (f) the application of written triage criteria that form the basis for bypassing non-trauma centre hospitals; (g) the presence of ongoing monitoring systems for trauma centres; and (h) the state-wide availability of trauma centres.

The integral steps in developing a regional trauma system are (37):

1. Basic Data

The first step is defining the magnitude of the problem in the area to be regionalised. This can be carried out using autopsy studies (38,39), preventable death studies (40), and/or regional trauma reviews (41). Out-of region experts should be recruited in order to provide objective assessments of the system in place.

2. Develop a Comprehensive Regional Plan

The regional plan should deal with patient care from the time of injury until the end of their rehabilitation. It should be based on guidelines from the American College of Surgeons (42,43,44) and have local surgeons heavily involved in planning and development.

The plan should address the following issues:
- Pre-hospital Care
- Air Transport
- Triage
- Trauma Centre Designation
- Quality Assurance
- Specialty Care Programs
- Research
- Rehabilitation
- Prevention and Public Education
- Disaster Planning

3. Identify Barriers to Change

By identifying barriers to changes prior to attempted implementation, a young system can develop strategies to overcome these changes. The major barriers to change are usually economic.

4. Develop a Management Structure

A lead agency must be identified and given formal, legal authority for trauma centre designation.

5. How to Implement the Plan

Once the plan has been developed, all regional hospitals should be encouraged to participate and undergo formal verification.

An "inclusive" approach to trauma system design has been adopted by trauma system planners (45). This approach is designed to improve the quality of care provided to injured patients by developing strategies for overcoming problems of access, cost and variation in the quality of services. Planning and implementing a system of trauma care is a huge undertaking (46). It requires intensive study, coordination and financial commitment. In the United States, the problem of access for patients without health insurance and those in rural areas have become paramount to the

"inclusive" system. These problems are constantly being investigated and commitment on the part of systems for the care of these patients are vital to the success of these systems in the future.

THE ELEMENTS OF A TRAUMA SYSTEM

A model trauma care system includes the basic concept of "inclusiveness". An inclusive system encompasses all aspects of trauma from prevention of injury until the patient returns to their pre-injury baseline level of function. The key elements of regionalised trauma systems are: (1) a lead public agency with legal authority to establish and enforce trauma system policy; (2) facility categorization; (3) trauma centre designation; and (4) the implementation of triage and transfer protocols which identify patients in need of transport to definitive care at a designated trauma care centre (47,48,49). Even though these elements are essential and common across all trauma systems, individual variations exist. These variations are present in the methods different communities use to design, implement and run their systems. These differences are profound in the area of the process of trauma centre designation (48). Bazzoli et al identify three key elements integral to trauma care regionalisation: pre-hospital care, organization of hospitals and inter-hospital transfer agreements (50). By assuring appropriate and timely inter-hospital transfers, patients can be appropriately treated in a system encompassing remote and rural areas (51).

The American Trauma Society (ATS) identifies four fundamental components necessary for trauma systems and eight key infrastructure elements that are critical to trauma system success (52):

Fundamental Components

- Injury Prevention
- Pre-hospital Care
- Acute Care Facilities
- Post-hospital Care

Key Infrastructure Elements

- Leadership
- Professional Resources
- Education and Advocacy
- Information Management
- Finances
- Research
- Technology
- Disaster Preparedness and Response - Conventional and Unconventional

Time-distance relationships between injured patients and definitive and appropriate care are vital to any trauma system design (53). Systems need to be created with geographic, time-transportation factors and maximum health delivery capabilities of a region in mind (54).

Another crucial element involved in maintaining an effective regionalised trauma system is quality improvement. Effective and continuous quality improvement programs depend upon concurrent monitoring of the events involved and surrounding the care of the trauma patient (21). The information for quality improvement programs is usually stored in a trauma databank, maintained either at the individual institutions within the system, or in a centralised databank for the entire system, state/province or country. Important elements to be evaluated include: facts related to the patient's injury event, injury severity, process of care and outcome.

Pre-hospital triage algorithms are integral to the optimal care for the injured patient. Injured patients need to be taken to the appropriate level facility that is prepared, properly staffed, and equipped to handle the trauma patient. Various schemes have been proposed for the pre-hospital triage of trauma patients. The most widely used is probably the American College of Surgeons Triage Algorithm (55,34). Triage schemes have been shown to be effective at decreasing trauma mortality (56,57,58). The algorithms outline strategies for transporting the seriously injured patient to an appropriate centre, bypassing lower level centres, which are often closer to the scene of the accident.

Trauma centres remain a key component in the systems approach to the acute care of the severely injured patient (59,60). Designation of these centres is integral to improving outcomes (36,61). By having designated centres committed to the resource allocation and care of injured patients, improvements in both morbidity and mortality have been demonstrated. However, the system encompasses all phases of care, from pre-hospital through acute care and rehabilitation. The creation and running of an effective system requires complete commitment from medical and allied health care professionals, as well as from regional health boards, governmental agencies and communities. Furthermore, even though the designation of trauma centres shifts more severely injured patients to designated hospitals (62), trauma centre care has been shown to significantly reduce length of stay and cost of care compared to injury severity matched patients transferred from a non-trauma facility (63). Patients directly transported to trauma centres also have less missed injuries than transferred patients (64). However, it has also been demonstrated that hospitals in remote areas that do not possess all elements necessary for the designation of trauma centres, can have similar, if not better, outcomes than those meeting criteria (65).

Surgical leadership is vital to maintaining an effective trauma system (66,67,68). The American

College of Surgeons Committee on Trauma emphasizes the role of the trauma surgeon in the design, implementation and running of a trauma system and trauma centre (34). The American Association for the Surgery of Trauma (AAST) expands on this and requires that a trauma surgeon be (69):

- Actively involved in the process of prehospital triage and treatment of trauma patients
- Thoroughly knowledgeable of the diagnostic options and treatment available in the emergency department and understands how to use them in the most appropriate and cost-effective manner
- Able to prioritise and coordinate the resuscitation and treatment of multiple serious injuries while coordinating care between multiple services and subspecialties
- Expert in the operative and nonoperative management of life-threatening and limb-threatening injuries
- Responsible for the comprehensive management of the injured patient in the critical care unit, including hemodynamic monitoring, ventilator management, nutrition and posttraumatic complications
- Integrally involved in the rehabilitation of the injured patient.
- Responsible for monitoring outcomes, identifying deficiencies in care when they exist, and correcting any identified deficiencies.
- Actively involved in trauma education, research and injury prevention.
- An advocate for the optimal care of trauma patients in public forums.

Another key element in the overall running of a trauma system is prevention (44). In fact, prevention is probably the single most effective way to decrease mortality and morbidity associated with injury.

Boyd appropriately points out that in order to design and implement an effective regional trauma system, focusing on one component of the subsystem will not be as effective as an overall and comprehensive view of the sequence of events as they affect the course and final outcome (53).

IN-HOSPITAL CARE - CHARACTERISTICS OF TRAUMA TREATMENT CENTRES

The categorisation of hospitals based on their ability to care for injured patients was first suggested by Youmans and Brose in 1970 (70). They conceptualised a classification system for hospitals treating injured patients in order to assure quality of care within a community. The initial classification system comprised: "major emergency facilities", "emergency facilities" and "provisional emergency facilities". These classifications later gave birth to level I, level II and level III trauma treatment centres.

Designated trauma centres have been shown to decrease mortality, complication rates, and length of hospital stay compared to non-trauma centres (71,72,73,74,75). Verification has also been shown to improve the process of care within trauma centres (76,77,78). An overview of a centre's role and requirements as part of a system of trauma care based on the ACS criteria for trauma centre designation is as follows:

Level I
Level I trauma centres are tertiary care facilities that are the focal point of a regionalised trauma system. These centres often, but not always occur in university hospitals. The facility must be capable of providing leadership and total care for every aspect of injury, from prevention to rehabilitation (59).

Level II
Level II trauma centres function in a similar capacity to level I centres, however, they do not have the extensive resources and facilities as level I centres. They are required to provide initial definitive trauma care to injured patients regardless of injury severity.

Level III
Level III centres usually occur in communities that do not have access to level I or II centres. These centres must have the capability to manage the initial care of the majority of injured patients and have transfer agreements and corridors set up for transfer of patients that exceed the hospitals resources and capabilities.

Level IV
Level IV centres are those centres treating and stabilizing injured patients in rural areas without other hospitals. They are the "de facto trauma centres" in these regions due to geographical location (59). They are responsible for providing Advanced Trauma Life Support care (55) in remote areas where no higher level of care is available prior to transfer to an advanced level centre.

The evaluation and management of severely injured patients requires significant institutional commitment and the commitment of skilled personnel (13). Recently, there has been much debate over the American College of Surgeons' requirements for minimal trauma centre volume in order for a centre to receive a designation (34). Numerous studies have been published over the last few years with conflicting results regarding the correlation between volume and outcome. Several studies have shown that volume has a positive correlation with survival (79,80,81,79,82,83,81), however others have demonstrated a lack of association (84,85,65,86,87). Guidelines for level I trauma centre verification require 1,200 admissions per year. Many centres in the US and Canada that cannot meet these requirements do, however, meet all other requirements for level I status.

THE IMPACT OF TRAUMA CARE REGIONALISATION

The initial fervour for trauma system implementation was backed by very few studies and lacked the large amounts of evidence that were to come over the years (88,89). However, since the late 1960s

there have been over thirty studies demonstrating a positive impact on survival in regionalised compared to non-regionalised trauma systems (Table 1). Furthermore, the lack of a trauma care system has also been repetitively shown to contribute to substandard care and outcomes (90,91,92,93,94). By centralising the care of severely injured patients in to a few highly specialized centres, as well as creating corridors for direct entry and easy exit from acute care, trauma systems significantly improve the outcome for injured patients (46,95,96,97,98,99,100,101,102,103) and change the pattern of preventable mortality from delays or inadequate interventions to postoperative care errors (104). Aggregated population-based evidence (61,71,72,73,105,106,107,108,109,110,111,112,113,114) has demonstrated a 15 to 20% improved survival rate for seriously injured patients following trauma system implementation (115).

Shackford (116) found that in the first year following establishment of a regionalised trauma system in San Diego County, severely injured patients (TS = 8) had a probability of survival (Ps) of 18% compared to injured patients treated at numerous centres throughout the US and Canada (117), and an actual survival of 29%. Many subsequently used this evidence in order to push healthcare systems and governments to establish organised systems of trauma care. San Diego County instituted a regionalised trauma system in 1984. Guss subsequently performed a before and after preventable death evaluation in the County (118,119) using the validated autopsy review methodology proposed by West (120). Preventable death evaluation involves the calculation of a preventable death rate (PDR), which is the proportion of all deaths judged to have been preventable if optimal care had been delivered (40). Guss found that by expert panel evaluation, 2 out of 211 deaths (1%) were preventable post regionalisation compared to 20 out of 177 (11.4%) pre-regionalisation ($p < 0.001$). Similar to the Orange County and San Francisco County patients, the decline in mortality post regionalisation was mostly attributed to a decline in mortality from non-central nervous system deaths.

Shackford studied the effect of regionalised trauma care on outcomes of "major trauma victims" in the first 5 months post-regionalisation and compared it to the period immediately prior to the implementation of a system in San Diego County using the medical audit committee technique for assessing optimal or suboptimal care (121). He found suboptimal care was rendered in 32% of cases prior to regionalisation, and that the implementation of a trauma system decreased the proportion to 4.2% (122). Preventable deaths occurred in 13.6% of fatalities prior to implementation, compared to 2.7% following system implementation. Shackford subsequently looked at a subset of severely injured trauma patients (Trauma Score of = 8) in the first year after trauma care regionalisation in San Diego County (121). He compared actual survival to predicted survival based on the Major Trauma Outcome Study (MTOS) (117). Following regionalisation, the probability of survival in blunt trauma patients was 18% compared to the 29% survival observed ($p<0.05$). In penetrating trauma, the probability and observed survivals were 8% and 20%, respectively ($p<0.05$).

Mullins evaluated the outcomes of trauma patients before and after institution of a regionalised trauma system. The risk of death in level I trauma centres improved following implementation of a regionalised system in the North Willamette region of Oregon between 1984 and 1991 (odds ratio = 0.65 post regionalisation) (109). The establishment of a regionalised trauma system also shifted the more seriously injured patients to the level I centres (123). Mullins then evaluated the influence of the implementation of a state-wide trauma system in Oregon on the location of hospitalisation and outcome of injured patients before and after regionalisation (110). In Oregon, following state-wide regionalisation, chances for an injured patient being admitted to a level I or II trauma centre increased and the chance of dying decreased.

A further study was done in order to attempt to control for temporal trends in advancements in medical and surgical care of injured patients (111). In this study injured patients in Oregon and Washington were compared before either state had a regionalised trauma system (1985-1988) as well as when only Oregon had a trauma system in place (1990-1993). Following trauma system implementation in Oregon, there was a significant risk reduction for death in patients with Injury Severity Scores > 15 (Odds Ratio = 0.8, CI = 0.70-0.91) compared to Washington. Pediatric mortality was also shown to be positively influenced by system implementation in Oregon, compared to Washington (107). Secular trends in trauma mortality are best adjusted by the types of studies that compare two systems over the same time period (124).

Kane evaluated the survival of seriously injured patients in Los Angeles County prior to (1982) and following (1984) implementation of a regionalised system of trauma care (108). There was an observed significant improvement in the adjusted odds of survival following regionalisation (odds ratio = 1.455, p-value = 0.048) compared to the period prior to the establishment of the system. Cayten reported on mortality following motor vehicle collisions in the Hudson Valley region of New York from 1987 to 1996

(125). There was also a significant decrease in motor vehicle collision mortality that was related and attributed to the establishment of a regionalised trauma system between 1990 and 1995.

Nathens evaluated the effect of trauma systems throughout the United States. He looked at data from states with organized trauma systems in place and compared them to those without regionalised trauma care (126). States that contained regionalised trauma systems (n=22) had a 9% lower crude mortality rate compared to those without regionalised care. After sub-analysis for motor-vehicle collisions, areas with organized trauma systems had a 17% reduction in mortality compared to those without systems.

Nathens also studied the effect of regionalised trauma care on motor vehicle crash mortality throughout the United States between 1979 and 1995 (127). He found that it took approximately 10 years following regionalisation of care to start to see a decline in mortality. By 15 years, mortality from motor vehicle collisions decreased by 8%. The 10-year interval between trauma system implementation and the improvement in outcomes was attributed to the necessary time for trauma system maturation, development of trauma triage protocols, inter-hospital transfer agreements, trauma centre organization, and ongoing quality assurance. These factors, however, were not assessed in this study and remain hypotheses.

Clark critically re-evaluated the aforementioned studies performed by Mullins (109), Cayten (125) and Nathens (127), which used data from the Fatality Analysis Reporting System (FARS), in order to test the accuracy of their results and assess the conclusions that were drawn regarding the effectiveness of trauma systems from these studies (128). He found that the positive impact of trauma system regionalisation was less convincing when all available data was displayed and potential confounding factors were assessed. Mortality following trauma was found to be decreasing throughout the United States and this contributed to the declining rates of mortality following injury. Clark's findings are controversial and have caused much debate (129). However, even if trauma systems do not impact on national mortality as much as some believe, they have and do definitely contribute to superior care for injured patients.

Jurkovich and Mock compared patients with serious injuries in three cities: Seattle (Washington), Monterrey (Mexico) and Kumasi (Ghana) (130,131,132). Seattle is considered to have the most advanced EMS service in the world, Monterrey has a basic EMS service and Kumasi has no EMS system. Major differences also obviously existed in hospital capabilities and socioeconomic factors. Overall survival for seriously injured patients were: Kumasi (36%), Monterrey (45%) and Seattle (65%). The increased survival was primarily attributed to decreased pre-hospital deaths, further highlighting the importance of the "system" in the outcome of seriously injured patients.

In July of 1998, a symposium was organized at the Skamania Lodge in Stevenson, Washington (133). The symposium was titled: "Trauma Systems - Evidence, Research, Action." The symposium was planned in order to assemble health care professionals from various disciplines to critically review the available evidence concerning trauma system effectiveness and was a huge success (134,135). Prior to the symposium, a comprehensive review of the literature was undertaken by the organizing committee and key articles concerning trauma system effectiveness were selected, summarized and sent to participants (136). The articles were then critiqued by the participants at the symposium and summarized in an important paper by Mann et al. in a supplement to the Journal of Trauma (10). Mann concluded that there was evidence supporting the effectiveness of regional trauma care systems in reducing in-hospital mortality. However, further outcome studies were required including studies based on 30-day post discharge mortality and the evaluation of morbidities.

Outcomes have also been shown to improve as time passes following establishment of a trauma system (137,138,139). As the system matures, mortality for severely injured patients declines. O'Keefe was able to show a positive survival advantage for injured patients with ISS = 16 over 10 years at a single level I trauma centre between 1986 and 1995 (140).

The effects of regionalisation in Canada have not been as extensively studied as the systems of trauma care in the United States. However, the impact of regionalisation on the outcome of trauma patients in the province of Quebec has been studied in depth over the last 15 years (58,93,138,141,142). Regionalisation of trauma care has been shown to significantly improve outcome for seriously injured patients in Quebec.

In the early years of trauma care regionalisation, designation of trauma care centres does not lead to increases in patient volume at designated trauma centres. Instead, there is a redistribution of patients, with the more severely injured patients being transported to the higher level centres (29,143). However, once a system becomes established and is running efficiently, outcomes improve (137) and proportions of trauma patients being transported to higher level centres increase (144,145). The increase in patients is usually secondary to the triage and transport of patients with low injury severity injuries. Pre-hospital care workers and dispatchers prefer

Table 1 . Studies Demonstrating Survival Benefit Following Trauma System Implementation

Study Author	Year	Region(s)	Data Source	Type of Patients	Number of Patients	Non-Regionalised Mortality	Regionalised Mortality	Change
Mullins RJ {158) }	1994	Four Counties in Portland, Oregon (pre- and post-regionalisation)	Hospital Discharge Data Analysis	Injured patients	25,145 pre 21,806 post	O.R. for death in Level I centres = 1.0		
Karsteadt LL (155)d}	1994	North Coast EMS Region of California (rural trauma system post-regionalisation compared to MTOS)	Regional Trauma Registry	Seriously injured patients	266 patients	MTOS Ps = 23.6%	Observed Survival = 20.3%	Z = -2.33 indicating lower mortality in study group compared to MTOS
Rutledge R {61)d}	1993	North Carolina Counties With Trauma Centres versus Counties Without Trauma Centres	State Medical Examiner's Database	Per capita trauma death rates .	309 (TC counties) 78 (non-TC counties)	5.0 deaths per 10,000 population	4.0 deaths per 10,000 population	Decrease in 1 death per 10,000 population
Hill DA {159)d}	1993	Royal Prince Alfred Hospital, Sydney, Australia (pre- versus post-in-hospital trauma integration)	Hospital-based Trauma Registry	Seriously injured patients	70 pre 51 post	28% crude mortality	11% crude mortality	17% improvement in crude mortality (no significant change in PDR)
Thoburn E (104)d}	1993	Hillsborough County, Florida (pre- and post-regionalisation)	Medical Audit - PDA	Non-CNS Trauma Deaths	452 pre 504 post	PDR = 23%	PDR = 7%	16% improvement in PDR
Kane G (108)d}	1992	Los Angeles County (pre- and post-regionalisation)	Hospital Chart Review	MVC with multiple serious injuries	658 pre 766 post	Odds of survival = 1.0	Odds of Survival = 1.455	45% increased odds of survival post-regionalisation
Rutledge R (72)d}	1992	North Carolina Counties With Trauma Centres versus CountiesWithout Trauma Centres	State Medical Examiner's Database	Per capita trauma death rates	309 (TC counties) 78 (non-TC counties)	5.0 deaths per 10,000 population	4.0 deaths per 10,000 population	Decrease in 1 death per 10,000 population
Champion HR (160)d}	1992	Washington Hospital Centre (pre- and post-trauma centre designation and system implementation)	Hospital-based Trauma Registry (TRISS)	Trauma with blunt mechanism	467 pre 214 post	Z = -2.17	Z = +1.78	4.34 more survivors per 100 patients post designation

Table 1 . Studies Demonstrating Survival Benefit Following Trauma System Implementation (continued)

Study Author	Year	Region(s)	Data Source	Type of Patients	Number of Patients	Non-Regionalised Mortality	Regionalised Mortality	Change
Smith JS ({73)d) }	1990	Western Pennsylvania and Maryland (Trauma Centres versus Non-Trauma Centres)	Hospital Discharge Data Analysis	Patients with femoral shaft fracture requiring operation	718 non-trauma centre 614 trauma centre	2.2% crude mortality	1.0% crude mortality	1.2% improvement in crude mortality (Non Significant)
Guss DA (119)d}	1989	San Diego County, California (pre- and post-regionalisation)	Medical Audit - PDA	All injury deaths	177 pre 211 post	PDR = 11.4%	PDR = 1%	10.4% improvement in PDR
Shackford SR {161)d}	1987	San Diego County (post-regionalisation compared to MTOS)	Medical Audit - PDA	Trauma Score = 8 .	189 patients	MTOS Ps = 18%	Observed Survival = 29%	11% reduction in mortality compared to MTOS
Shackford SR {122)d}	1986	San Diego County (pre- and post-regionalisation)	Medical Audit - PDA	Seriously injured patients	591 pre 1366 post	PDR = 13.6%	PDR = 2.7%	10.9% improvement in PDR
Clemmer TP (71)d}	1985	Salt Lake County, Utah (level I centres vs. community hospitals)	Hospital Discharge Data Analysis	CRAMS = 6	56 LI 24 community	46% crude mortality	75% crude mortality	29% improvement in crude mortality
Ornato JP (113)d}	1985	Nebraska (pre- and post-regionalisation)	Dept of Health Database	All injury deaths	474 pre 349 post			23.9% improvement in crude mortality
Cales RH (162)d}	1984	Orange County, California (pre- and post-regionalisation)	Medical Audit - PDA	MVC Deaths	58 pre 60 post	PDR = 34%	PDR = 15%	20% improvement in PDR
Alexander RH (106)d}	1984	Florida (comparison of counties with level I equivalent hospitals to those without)	Highway Patrol Database	Mileage Population Death Index (MPDI)	467 pre 214 post	MPDI = 24.5	MPDI = 0.82	Reduction in MPDI = 23.7
West JG (163)d}	1983	Orange County, California (pre- and post-regionalisation)	Medical Audit - PDA	Non-CNS MVC TC Deaths	21 pre 23 post	PDR = 71%	PDR = 9%	62% improvement in PDR
West JG (164)d}	1979	Orange County, California (non-regionalised) vs. San Francisco County, California (regionalised)	Medical Audit - PDA	Non-CNS MVC related deaths	30 non-regionalised 16 regionalised	PDR = 73%	PDR = 6%	67% improvement in PDR

Table 1. Studies Demonstrating Survival Benefit Following Trauma System Implementation (continued)

Study Author	Year	Region(s)	Data Source	Type of Patients	Number of Patients	Non-Regionalised Mortality	Regionalised Mortality	Change
Abernathy JH (151)d}	2002	Level I TC, University of Alabama at Birmingham (pre- Injured patients 1718 post and post- regionalisation)	Injured Patients Trauma Registry	Hospital-based	1306 pre 1718 post	5.9% crude mortality	3.8% crude mortality	Adjusted Odds of death post regionalisation = 0.48
Nathens AB (152)d}	2000	United States (22 regionalised statesversus non- regionalised states)	National Vital Statistics Database	All injury deaths	67,429 deaths	29.2/100,000	26.5/100,000	9% lower crude injury mortality rate in regionalize state
Nathens AB (127)d}	2000	United States (comparison between pre- regionalisation andpost- regionalisation in 22 states with trauma systems)	Fatality Analysis Reporting System	MVC mortality	439,195 deaths	16.9 deaths per 100,000 person-years	14.3 deaths per 100,000 person-.years	MRR = 0.87 (pre- vs post-regionalisation)
Özgüc H (153)d}	2000	Uldag University Medical School, Bursa, Turkey (pre- versus post-in-hospital trauma integration)	Medical Records Review	Seriously injured patients	242 pre 137 post	32.5% crude mortality	23.3% crude mortality	9.2% improvement in crude mortality
Nathens AB (154)d}	1999	United States (comparison between pre-regionalisation and post-regionalisation in 22 states with trauma systems)	Fatality Analysis Reporting System	MVC mortality		17.3 deaths per 100,000 person-years	14.2 deaths per 100,000 person-years	Adjusted Mortality Rate Ratio = 0.91 (pre- vs. post-regionalisation)
Sampalis JS (138)d}	1999	Montreal and Quebec City, Quebec (pre- and post-regionalisation)	Regional Trauma Database Analysis	Seriously injured patients	1,884 pre 2,107 post	51.8% crude mortality	17.7% crude mortality	34.1%% improvement in crude mortality (O.R.=0.147 post regionalisation)
Mullins RJ (111)d}	1998	Oregon (pre- and post-regionalisation) and Washington State (non-regionalised)	Hospital Discharge Data Analysis	Seriously injured patients	30,757 pre (OR + WA) 11,879 post (OR only)	R.R. for death = 1.0	R.R. for death = 0.91	9% decreased risk of death post regionalisation
Hulka F (107)d}	1997	Oregon (regionalisation) and Washington State (non-regionalised)	Hospital Discharge Data Analysis	Seriously injured children (<19 years)	12,991 non-regionalised 8,981 regionalised	O.R. for death = 1.0	O.R. for death = 0.68	32% adjusted risk-reduction for death in regionalised area
Mullins RJ (155)d}	1997	Oregon (pre- and post-regionalisation) and Washington State (non-regionalize state)	Hospital Discharge Data Analysis	Seriously injured patients	10,496 pre 10,629 post	O.R. for death = 0.92	O.R. for death = 0.80	8% decreased odds of death post regionalisation

Table 1. Studies Demonstrating Survival Benefit Following Trauma System Implementation (continued)

Study Author	Year	Region(s)	Data Source	Type of Patients	Number of Patients	Non-Regionalised Mortality	Regionalised Mortality	Change
Rogers FB (114)d}	1997	San Diego County (urban regionalised trauma system) compared to Vermont (rural non-regionalised system)	State Medical Examiner's Database and Autopsy Database	On-scene deaths .	103 non-regionalised 248 regionalised	72% crude mortality	40.5% crude mortality	Rural/non-regionalised have significantly higher proportion of on-scene deaths
Nichol J (156)d}	1997	North Staffordshire Royal Infirmary and 5 district general hospitals in North West Midlands (regionalised) vs. Lancashire and Humberside, England (non-regionalised)	Medical Records Review	Seriously injured patients	1503 non-regionalised 1143 regionalised	45% crude mortality	43% crude mortality	No difference in either crude or adjusted mortality
Mullins RJ (110)d}	1996	State of Oregon (pre- and post-regionalisation)	Hospital Discharge Data Analysis	Injured patients	14,694 pre 13,654 post	O.R. for death = 1.0	O.R. for death = 0.82	18% decreased odds of death post regionalisation
Sampalis JS (142)d}	1995	Montreal, Quebec (pre- and post-trauma centre designation)	Regional Trauma Database Analysis	Seriously injured patients	158 pre 288 post	20% crude mortality	10% crude mortality	Mortality pre-regionalisation significantly greater than MTOS, post-regionalisation no change
Stewart TC (157)d}	1995	Victoria Hospital, London, Ontario (pre- and post-regionalisation)	Hospital-based Trauma Registry (TRISS)	MVC patients, ISS>12	156 pre 189 post	Z = -0.40	Z = +0.72	6 more survivors per 100 death post regionalisation patients post regionalisation
Waters JM (165)d}	1973	Jacksonville, Florida (pre- and post-regionalisation)		All traffic accidents	16,035 pre 22,494 post	8.4 deaths per 1000 accidents	5.2 deaths per 1000 accidents	38% reduction in mortality from traffic accidents
Boyd DR (105)d}	1973	State of Illinois (pre- and post-regionalisation)	Highway Death Rate Analysis	All highway injuries		2.8% mortality rate	2.1% mortality rate	Decline in highway fatalities of 8%

*****Based on the % reference according to each individual patient's age, sex, and height. +This parameter is not a score but the actual dose converted into beclomethasone units (mcg) d Statistically significant p-values accounting for non-independence *PDA = Preventable Death Analysis; PDR = Preventable Death Rate; MVC = Motor Vehicle Collision; TC = Trauma Centre; O.R. = Odds Ratio; R.R. = Relative Risk; MRR = Mortality Rate Ratio; Mileage Population Death Index (MPDI) = average death rate per one hundred million miles driven divided by the population at risk x 105; MTOS = Major Trauma Outcome Study (The MTOS is an outcome database of approximately 160,000 trauma patients treated between 1982 and 1989 at 139 hospitals throughout the United States and Canada); Ps = Probability of Survival (based on MTOS data)*

to err on the side of over-triage in order not to miss significant occult injuries (146,147). Furthermore, triage algorithms are designed to over-triage less severely injured patients (43,148,149,150). These factors contribute significantly to the high costs of running a level I trauma centre (48).

CONCLUSION

Regionalisation of trauma care improves outcome for injured patients by utilizing a systematic approach to the care of the injured patient. This approach encompasses all phases of injury from prevention to rehabilitation. A systematic approach to the care of the trauma patient is based on cooperation between pre-hospital emergency medical services, hospitals of all levels, rehabilitation facilities and local, regional, state-wide/provincial and national organizations. By pooling resources and emphasizing teamwork and cooperation, trauma systems have changed the face of trauma patient care, significantly decreased morbidity and mortality secondary to injury and set a benchmark for the regionalised approach to patient management for other areas of healthcare.

REFERENCES

1. Health Reports 1995. Statistics Canada, 1995.
2. Mortality. Summary List of Causes, Vital Statistics. Health Division, Vital Statistics and Health Status Section. Ottawa: Statistics Canada, 1986.
3. Mortality. Summary List of Causes, Vital Statistics. Health Division, Vital Statistics and Health Status Section. Ottawa: Statistics Canada, 1987.
4. Mortality. Summary List of Causes, Vital Statistics. Health Division, Vital Statistics and Health Status Section. Ottawa: Statistics Canada, 1988.
5. Gupta Kj, Parr MJA, and Nolan JP. Epidemiology of trauma - The role of the anesthesiologist. Anesth Clin N Am 1999;17:17-32.
6. MacKenzie EJ, Siegel JH, Shapiro S, Moody M, and Smith RT. Functional recovery and medical costs of trauma: An analysis by type and severity of injury. J Trauma 1988;28:281-297.
7. Rivara FP, Grossman DC, and Cummings P. Injury Prevention (first of two parts). NEJM 1997;337:543-548.
8. Rivara FP, Grossman DC, and Cummings P. Injury prevention (second of two parts). NEJM 1997;337:613-618.
9. Elliott DC and Rodriguez A. Cost effectiveness in surgery. Surg Clin N Am 1996;76:47-62.
10. Mann NC, Mullins RJ, MacKenzie EJ, Jurkovich GJ, and Mock CN. Systematic review of published evidence regarding trauma system effectiveness. J Trauma 1999;47:S25-S33.
11. Harlan LC, Harlan WR, and Parsons PE. The economic impact of injuries: A major source of medical costs. Am J Pub Health 1990;80:453.
12. Miller TR and Levy DT. The effect of regional trauma care systems on costs. Arch Surg 1995;130:188-193.
13. Eastman AB, Lewis FR, Champion HR, and Mattox KL. Regional Trauma system design: critical concepts. Am J Surg 1987;154:79-87.
14. Eastman AB. Blood in our streets. The status and evolution of trauma care systems. Arch Surg 1992;127:677-681.
15. Eastman AB, Schwab CW, Annest JL, Aprahamian C, Beachley M, Browner B, and et al. Position paper on trauma care systems. J Trauma 1992;32:127-129.
16. Boyd DR. Trauma-A controllable disease in the 1980's (Fourth Annual Stone Lecture, American Trauma Society). J Trauma 1980;20:14-24.
17. Eggold R. Trauma care regionalization: a necessity (editorial. J Trauma 1983;23:260-262.
18. Farmer TA. Trauma center organization and operation in academic health centers. Am J Emerg Med 1985;3:244-246.
19. Trunkey DD. Trauma. Scientific American 1983;249:28-35.
20. McMurtry RY, Nelson WR, and de la Roche MR. Current concepts in trauma: 1. Principles and directions for development. CMAJ 1989;141:529-533.
21. Rutledge R. The goals, development, and use of trauma registries and trauma data sources in decision making in injury. Surg Clin North Am 1995;75:305-326.
22. Conference on Injury in America. Public Health Rep 1987;102:577.
23. Waller JA. Emergency care for fatalities from injury and illness in the nonhighway setting. J Trauma 1973;13:54-60.
24. Baker CC, Oppenheimer L, Stephens B, Lewis FR, and Trunkey DD. Epidemiology of trauma deaths. Am J Surg 1980;140:144-150.
25. Trunkey DD. Initial treatment of patients with extensive trauma. NEJM 1991;324:1259-1263.
26. Maull KI and Esposito TJ. Trauma System Design. Chapter 4. Trauma System Design. Mattox KL, Feliciano DV, and Moore EE. Fourth Edition ed. New York: McGraw-Hill Companies Inc., 2000.
27. Beebe GW and DeBakey ME. Battle Casualties. Springfiel, Illinois: Charles C Thomas, 1952.
28. Zollinger RW. Traffic injuries - A surgical problem. Arch Surg 1955;70:694-700.
29. Cales RH, Anderson PG, and Heilig RW. Utilization of medical care in Orange County: the effect of implementation of a regional trauma system. Ann Emerg Med 1985;14:853-858.
30. Trunkey DD. Trauma care systems. Emerg Med Clin N Am 1984;2:913-922.
31. Robertson LS. Injury Epidemilogy. second edition ed. New York: Oxford University Press, 1998.
32. Olson CJ, Brand D, Mullins RJ, Harrahill M, and Trunkey DD. Time to death of hospitalized injured patients as a measure of quality of care. J Trauma 2003;55:45-52.
33. Dailey JT, Teter H, and Cowley RA. Trauma center closures: a National assessment. J Trauma 1992;33:539-547.
34. American College of Surgeons, Committee on Trauma. Resources for optimal care of the injured patient: 1999. American College of Surgeons, 1998.
35. Trunkey DD. Regionalization of trauma care. Topics Emerg Med 1981;3:91-96.
36. Aprahamian C, Wolferth CC, Darin JC, McMahon J, and Weitzel-DeVeas. Status of trauma center designation. J Trauma 1989;29:566-570.
37. West JG, Williams MJ, Trunkey DD, and Wolferth CC. Trauma systems. Current status - future challenges. JAMA 1988;259:3597-3600.
38. West JG. An autopsy method for evaluating trauma care. J Trauma 1981;21:32-34.
39. West JG. Validation of autopsy method for evaluating trauma care. Arch Surg 1982;117:1033-1035.
40. MacKenzie EJ, Steinwachs DM, Bone LR, Floccare DJ, Ramzy AI, and The preventable death study group. Inter-rater reliability of perventable death judgments. J Trauma 1992;33:292-303.
41. Cales RH. Medical Audit. Chapter 9. Medical Audit. Cales RH and Heilig RW Jr Rockville, MD: Aspen Systems Corporation, 1986.

42. American College of Emergency Physicians. Guidelines for trauma care systems. Ann Emerg Med 1987;16:459.
43. American College of Emergency Physicians. Guidelines for transfer of patients. Ann Emerg Med 1985;14:1221.
44. American College of Emergency Physicians. Trauma care systems development, evaluation, and funding. Ann Emerg Med 1999;34:308.
45. Hackey RB. The politics of trauma system development. J Trauma 1995;39:1045-1053.
46. Esposito TJ, Nania J, and Maier RV. State trauma system evaluation: a unique and comprehensive approach. Ann Emerg Med 1992;21:351-357.
47. Boyd DR and Cowley RA. Comprehensive regional trauma/emergency medical services (EMS) delivery systems. World J Surg 1983;7:149-157.
48. Goldfarb MG, Bazzoli GJ, and Coffey RM. Trauma systems and the costs of trauma care. HSR 1996;31:71.
49. Schwab W, Frankel HL, Rotondo MF, Gares DA, Robison EA, Haskell RM, and et al. The impact of a true partnership between a university level I trauma center and a level II trauma center on patient transfer practices. J Trauma 1998;44:815-820.
50. Bazzoli GJ, Madura KJ, Cooper GF, MacKenzie EJ, and Maier RV. Progress in the development of trauma systems in the United States: results of a national survey. JAMA 1995;273:395-401.
51. Cone JB. Tertiary trauma care in a rural state. Am J Surg 1990;160:652-658.
52. American Trauma Society Supported by the U.S.Department of Transportation, National Highway Traffic Safety Administration. Trauma System Agenda for the Future. 2002.
53. Boyd DR, Pizzano WA, Romano TL, Stiegmann GV, and Nyhus LM. Regionalization of trauma patient care: the Illinois experience. Surg Annu 1975;7:25-52.
54. Waller JA. Emergency health services in areas of low population density. JAMA 1969;207:2255-2258.
55. American College of Surgeons. Advanced Trauma Life Support for Doctors (ATLS) Course Manual. Sixth Edition ed. Chicago, IL: American College of Surgeons, 1997.
56. Lyle DM, Thomson PC, Deane SA, and Coulon LA. Regionalization of trauma services in Western Sydney: predicted effect on ambulance and hospital utilization rates. Aust NZ J Surg 1991;61:589-596.
57. Lyle DM, Thompon PC, Coulon LA, Berry G, and Kim CM. Trauma triage in Western Sydney: results of a pilot study. Aust NZ J Surg 1990;60:953-958.
58. Sampalis JS, Denis R, Frechette P, Brown R, Fleiszer D, and Mulder D. Direct transport to tertiary trauma centers versus transfer from lower level facilities: impact on mortality and morbidity among patients with major trauma. J Trauma 1997;43:288-296.
59. Arroyo JS and Crosby LA. Basic rescue and resuscitation. Trauma System concept in the United States. Clinical Orthopaedics and Related Research 1995;318:11-16.
60. Thompson CT. The emergency physician, the trauma surgeon, and the trauma center. Ann Emerg Med 1983;12:235-237.
61. Rutledge R, Fakhry SM, Meyer M, Sheldon GF, and Baker CC. An analysis of the association of trauma centers with per capita hospitalizations and death rates from injury. Ann Surg 1993;218:521-524.
62. Hammond JS and Breckenridge MB. Longitudinal analysis of the impact of a level 1 trauma center designation at a university hospital. J Am Coll Surg 1999;188:217-224.
63. Schwartz RJ, Jacobs LM, and Yaezel D. Impact of pre-trauma center care on length of stay and hospital charges. J Trauma 1989;29:1611-1615.
64. Robertson R, Mattox R, Collins T, Parker-Mills C, Eidt J, and Cone J. Missed injuries in a rural area trauma center. Am J Surg 1996;172:567-568.
65. Allen DM, Hicks G, and Bota GW. Outcomes after severe trauma at a Northern Canadian regional trauma centre. Can J Surg 1998;41:53-58.
66. Lowe DK. Trauma system development: the critical need for regional needs assessments. J Trauma 1999;47:S106-S107.
67. Moore EE. Trauma sysrems, trauma centers, and trauma surgeons-Opportunity in managed competition. J Trauma 1995;39:1-11.
68. Mulder DS. Specialization in surgery - Implications for trauma-related disciplines. Bull Am Coll Surg 1988;73:15.
69. Hoyt DB, Moore EE, Shackford SR, Holcroft JW, and Jurkovich GJ. Trauma surgeon's leadership role in the development of trauma systems (abstract). J Trauma 1999;46:1142.
70. Youmans RL and Brose RA. A basic for classifying hospital emergency services. JAMA 1970;213:1647-1651.
71. Clemmer TP, Orme JF, Thomas FO, and Brooks KA. Outcome of critically injured patients treated at level I trauma centers versus full-service community hospitals. Crit Care Med 1985;13:861-863.
72. Rutledge R, Messick J, Baker CC, Rhyne S, Butts J, Meyer A, and Ricketts T. Multivariate population-based analysis of the association of county trauma centers with per capita county trauma death rates. J Trauma 1992;33:29-38.
73. Smith JS, Martin LF, Young WW, and Macioce DP. Do trauma centers improve outcome over non-trauma centers: the evaluation of regional trauma care using discharge abstract data and patient management categories. J Trauma 1990;30:1533-1538.
74. Piontek FA, Coscia R, Marselle CS, Korn RL, and Zarling EJ. Impact of American college of surgeons verification on trauma outcomes. J Trauma 2003;54:1041-1047.
75. DiRusso S, Holly C, Kamath R, Cuff S, Cuff S, Sullivan T, Scharf H, and et al. Preparation and achievement of American college of surgeons level I trauma verification raises hospital performance and improves patient outcome. J Trauma 2001;51:294-300.
76. Simons RK, Kasic S, Kirkpatrick AW, Vertesi L, Phang T, and Appleton L. Relative improtance of designation and accreditation of trauma centers during evolution of a regional trauma system (abstract). 2001.
77. Simons R, Kasic S, Kirkpatrick A, Vertesi L, Phang T, and Appleton L. Relative importance of designation and accreditation of trauma centers during evolution of a regional trauma system. J Trauma 2002;52:827-834.
78. Ehrlich PF, Rockwell S, Kincaid S, and Mucha P. American College of Surgeons, Committee on Trauma verification review: does it really make a difference? J Trauma 2002;53:811-816.
79. Pasquale MD, Peitzman AB, Bednarski J, and Wasser TE. Outcome analysis of Pennsylvania trauma centers: Factors predictive of nonsurvival in seriously injured patients. J Trauma 2001;50:465-474.
80. Nathens AB, Jurkovich GJ, Maier RV, Grossman DC, MacKenzie EJ, Moore M, and Rivara FP. Relationship between trauma center volume and outcomes. JAMA 2001;285:1164-1171.
81. Smith RF, Frateschi L, Sloan EP, Campbell L, Kreig R, Edwards LC, and Barrett JA. The impact of volume on outcome in seriously injured trauma patients: two years' experience of the Chicago trauma system. J Trauma 1990;30:1066-1076.
82. Konvolinka CW, Copes WS, and Sacco WJ. Institution and per-surgeon volume versus survival outcome in Pennsylvania's trauma centers. Am J Surg 1995;170:333-340.
83. Pasquale MD, Peitzman AB, and Wasser TE. Outcome analysis of level 1 and 2 trauma centers (TC's). J Trauma 1998;45:189.
84. London JA and Batistella FD. Is there a relationship between trauma center volume and mortality? J Trauma 2003;54:16-25.

85. Cooper A, Hannan EL, Bessey PQ, Farrell LS, Cayten CG, and Mottley L. An examination of the volume-mortality relationship for New York State trauma centers. J Trauma 2000;48:16-24.
86. Tepas JJ, Patel JC, DiScala C, Wears RL, and Veldenz HC. Relationship of trauma patient volume to outcome experience. Can a relationship be defined? J Trauma 1998;44:827-831.
87. Richardson JD, Schmieg R, and Boaz P. Impact of trauma attending surgeon case volume on outcome: is more better? J Trauma 1998;44:266-272.
88. Hoffer EP. Emergency Medical Services, 1979. NEJM 1979;301:1118-1121.
89. Trafton PG and Trunkey DD. Regionalization of trauma care (letter). NEJM 1980;302:1094-1095.
90. Van Wagoner FH. Died in hospital: a three year study of deaths following trauma. J Trauma 1961;1:401-408.
91. Wright CS, McMurtry RY, Hoyle M, and Pickard J. Preventable deaths in mulitple trauma: review of deaths at Sunnybrooke Medical Centre Trauma Unit. CJS 1983;26:20-23.
92. Kreis DJ, Plasencia G, Augenstein D, Davis JH, Echenique M, Vopal J, and et al. Preventable trauma deaths: Dade County, Florida. J Trauma 1986;26:649-654.
93. Sampalis JS, Lavoie A, Williams JI, Mulder DS, and Kalina M. Standardized mortality ratio analysis on a sample of severely injured patients form a large Canadian city without regionalized trauma care. J Trauma 1992;33:205-212.
94. Lowe DK, Gately HL, Goss JR, Frey CF, and Peterson CG. Patterns of death, complication and errors in the management of motor vehicle accident victims: implications for a regional system of trauma care. J Trauma 1985;28:503-509.
95. Urbach DR, Bell CM, and Austin PC. Differences in operative mortality between high-and low-volume hospitals in Ontario for 5 major surgical procedures: estimating the number of lives potentially saved through regionalization. CMAJ 2003;168:1409-1414.
96. Gross M. Allocation of scarce resources: assessment and impact. CJS 2002;45:8.
97. Brackstone M, Doig GS, and Girotti MJ. Surgical case costing: trauma is underfunded according to resource intensity wieghts. CJS 2002;45:57-62.
98. Birkmeyer JD, Siewers AE, Finlayson EVA, Stukel TA, Lucas FL, Batista I, and et al. Hospital volume and surgical volume in the United States. NEJM 2002;346:1128-1137.
99. Dudley RA, Johansen KL, Brand R, Rennie DJ, and Milstein A. Selective referral to high-volume hospitals : estimating potentially avoidable deaths. JAMA 2000;283:1159-1166.
100. Hemmelgarn BR, Ghali WA, and Quan H. A case study of hospital closure and centralization of coronary vascular procedures. CMAJ 2001;164:1431-1435.
101. Sibbald B. Move to close low-volume surgical sites gaining spee. CMAJ 2002;167:177.
102. Wexler MJ. More procedures, better quality of care? Is there a case for regionalization of pancreatic resection for neoplasm? CMAJ 1999;160:671-673.
103. Simunovic M, To T, Theriault M, and Langer B. Relation between hospital surgical volume and outcome for pancreatic resection for neoplasm in a publicly funded healthcare system. CMAJ 1999;160:643-648.
104. Thoburn E, Norris P, Flores R, Goode S, Rodriguez E, Adams V, Campbell S, Albrink M, and Rosemurgy A. System care improves trauma outcome: Patient care errors dominate reduced preventable death rate. J Emerg Med 1993;11:135-139.
105. Boyd DR, Mains KD, and Flashner BA. A systems approach to statewide emergency medical care. J Trauma 1973;13:276-284.
106. Alexander RH, Pons PT, Krischer J, and Hunt P. The effect of advanced life support and sophisticated hospital systems on motor vehicle mortality. J Trauma 1984;24:486-490.
107. Hulka F, Mullins RJ, Mann NC, Hedges JR, Rowland D, Worrall WH, and et al. Influence of a statewide trauma system on pediatric hospitalization and outcome. J Trauma 1997;42:514-519.
108. Kane G, Wheeler NC, Cook S, Englehardt R, Pavey B, Green K, and et al. Impact of the Los Angeles County trauma system on the survival of seriously injured trauma patients. J Trauma 1992;32:576-583.
109. Mullins RJ, Veum-Stone J, Helfand M, Zimmer-Genbeck M, Hedges JR, Southard PA, and Trunkey DD. Outcome of hospitalized injured patients after institution of a trauma system in an urban area. JAMA 1994;271:1919-1924.
110. Mullins RJ, Veum-Stone J, Hedges JR, Zimmer-Gembeck MJ, Mann C, Southard PA, and et al. Influence of a statewide trauma system on location of hospitalization and outcome of injured patients. J Trauma 1996;40:536-546.
111. Mullins RJ, Mann NC, Hedges JR, Worrall W, and Jurkovich GJ. Preferential benefit from implementation of a statewide trauma system in one of two adjacent states. J Trauma 1998;44:609-617.
112. Mullner R and Goldberg J. An evaluation of the Illinois trauma system. Med Care 1978;16:140-151.
113. Ornato JP, Craren EJ, Nelson NM, and Kimball KF. Impact of improved emergency medical services and emergency trauma care on the reduction in mortality from trauma. J Trauma 1985;25:575-579.
114. Rogers FB, Shackford SR, Hoyt DB, Camp L, Osler TM, Mackersie RC, and Davis JW. Trauma deaths in a mature urban vs rural trauma system. Arch Surg 1997;132:376-382.
115. Mullins RJ and Mann NC. Population-based research assessing the effectiveness of trauma systems. J Trauma 1999;47:S59-S66.
116. Shakford SR, Mackersie RC, Hoyt DB, Baxt WG, Eastman AB, Hammill FN, Knotts B, and Virgilio RW. Impact of a trauma system on outcome of severely injured patients. Arch Surg 1987;122:523-527.
117. Champion HR, Copes WS, Sacco WJ, Lawnick MM, Keast SL, Bain LW, and et al. The major trauma outcome study: Establishing National norms for trauma care. J Trauma 1990;30:1356-1365.
118. Neuman TS, Bockman MA, Moody P, Dunford JV, Griffith LD, Guber SL, Guss DA, and Baxt WG. An autopsy study of traumatic deaths. Am J Surg 1982;144:722-727.
119. Guss DA, Meyer FT, Neuman TS, Baxt WG, Dunford JV, Griffith LD, and Guber SL. The impact of a regionalized trauma system on trauma care in San Diego County. Ann Emerg Med 1989;18:1141-1145.
120. West JG. An autopsy method for evaluating trauma care. J Trauma 1981;21:32-34.
121. Shackford SR, Hollingsworth-Fridlund P, McArdle M, and Eastman AB. Assuring quality in a trauma system - the medical audit committee: composition, cost and results. J Trauma 1987;27:866-875.
122. Shackford SR, Hollingsworth-Fridlund P, Cooper GF, and Eastman AB. The effect of regionalization upon the quality of trauma care as assessed by concurrent audit before and after institution of a trauma system: a preliminary report. J Trauma 1986;26:812-820.
123. Mann NC, Hedges JR, Sandoval R, Worrall W, Zechnich AD, Jurkovich GJ, and Mullins RJ. Trauma system impact on admission site- A comparison of two states. J Trauma 1999;46:631-637.
124. Hedges JR. Summary of the discussion: what have we learned about population-based investigations. J Trauma 1999;47:S67-S68.
125. Cayten CG, Quervalu I, and Agarwal N. Fatality analysis reporting system demonstrates association between trauma system initiatives and decreasing death rates. J Trauma 1999;46:751-756.
126. Nathens AB, Jurkovich G, Cummings P, Rivara FP, and Maier RV. Effectiveness of state trauma systems - A National evaluation. J Trauma 1999;47:221.
127. Nathens AB, Jurkovich GJ, Cummings P, Rivara FP, and Maier RV. The effect of organized systems of trauma care on motor vehicle crash mortality. JAMA 2000;283:1990-1994.

128. Clark DE. Trauma system evaluation using the fatality analysis reporting system. J Trauma 2003;54:1199-1204.
129. Mullins RJ. Editiorial comment - Trauma systems evaluation using the fatality reporting system. J Trauma 2003;54:1204.
130. Jurkovich GJ and Mock C. Systematic review of trauma system effectiveness based on registry comparisons. J Trauma 1999;47:S46-S55.
131. Mock C, Jurkovich GJ, nii-Amon-Kotei D, Arreola-Risa C, and Maier R. Trauma mortality patterns in three Nations at different economic levels - Implications for global trauma system development. J Trauma 1997;43:193.
132. Mock CN, Adzotor KE, Conklin E, Denno DM, and Jurkovich GJ. Trauma outcomes in the rural developing world: comparison with an urban level I trauma center. J Trauma 1993;35:518-523.
133. Mullins RJ and Mann NC. Introduction to the academic symposium to evaluate evidence regarding the efficacy of trauma systems. J Trauma 1999;47:S3-S7.
134. Hunt RC. Reflections on the Skamania conference: An emergency physician's perspective. J Trauma 1999;47:S108.
135. Atweh NA. Toward the all-inclusive trauma system. J Trauma 1999;47:S109.
136. Mullins RJ and Mann NC. Development of a systematic review of published evidence regarding the efficacy of trauma systems. J Surg Outcomes 1998;1:53.
137. Peitzman AB, Courcoulas AP, Stinson C, Udekwu AO, Billiar TR, and Harbrecht BG. Trauma center maturation - Quantification of process and outcome. Ann Surg 1999;230:87-94.
138. Sampalis JS, Denis R, Lavoie A, Frechette P, Boukas S, Nikolis A, and et al. Trauma care regionalization: A process-outcome evaluation. J Trauma 1999;46:565-581.
139. Barquist E, Pizzutiello M, Tian L, Cox C, and Bessey P. Effect of trauma system maturation on mortality rates in patients with blunt injuries in the Finger Lakes Region of New York State. J Trauma 2000;49:63-70.
140. O'Keefe GE, Jurkovich GJ, Copass M, and Maier RV. Ten-year trend in survival and resource utilization at a level I trauma center. Ann Surg 1999;229:409-415.
141. Sampalis JS, Lavoie A, Williams JI, Mulder DS, and Kalina M. Impact of on-site care, pre-hospital time, and level of in-hospital care on survival in severely injured patients. J Trauma 1993;34:252-261.
142. Sampalis JS, Lavoie A, Boukas S, Tamim H, Nikolis A, Frechette P, Brown R, Fleiszer D, Denis R, Bergeron E, and Mulder D. Trauma center designation: initial impact on trauma-related mortality. J Trauma 1995;39:232-239.
143. Goldberg J, Levy PS, Gelfand HM, Mullner R, Iverson N, Lemeshow S, and Rothrock J. Factors affecting trauma center utilization in Illinois. Med Care 1981;19:547-565.
144. Liberman M, Mulder D, and Sampalis JS. Increasing Volume of Low Injury Severity Elderly Patients at Level I Trauma Centers: Is There a Need for Triage Modification? Can J Surg. 2003 46:446-52.
145. Hedges JR, Mullins RJ, Zimmer-Gembeck M, Helfand M, and Southard P. Oregon trauma system: change in intial admission site and post-admission transfer of injured patients. Acad Emerg Med 1994;1:218-226.
146. Esposito TJ, Offner PJ, Jurkovich GJ, Griffith J, and Maier RV. Do prehospital trauma center triage criteria identify major trauma victims? Arch Surg 1995;130:171-176.
147. Báez AA, Lane PL, and Sorondo B. System compliance with out-ofhospital trauma triage criteria. J Trauma 2003;54:344-351.
148. American College of Surgeons. Advanced Trauma Life Support for Doctors (ATLS). Sixth Edition ed. Chicago, IL: American College of Surgeons, 1997.
149. Champion HR, Sacco WJ, Hannan DS, Lepper RL, Atzinger ES, CopesWS, and Prall RH. Assessment of injury severity: the triage index. Crit Care Med 1980;8:201-208.
150. Gormican SP. CRAMS scale: field triage of trauma victims. Ann Emerg Med 1982;11:132-135.
151. Abernathy JH III, McGwin G Jr, Acker JE III, and Rue LW III. Impact of a voluntary trauma system on mortality, length of stay, and cost at a level I trauma center. Am Surgeon 2002;68:182-192.
152. Nathens AB, Jurkovich GJ, Rivara FP, and Maier RV. Effectiveness of state trauma systems in reducing injury-related mortality: A National evaluation. J Trauma 2000;48:25-31.
153. Ozguc H, Kaya E, Yunuk O, Armagan E, and Tokyay R. Outcome of major trauma in a Turkish university hospital: did integrated approach make a difference? Eur J Emerg Med 2000;7:183-188.
154. Nathens AB, Jurkovich G, Cummings P, Rivara FP, and Maier RV. Effectiveness of state trauma systems - A National evaluation (abstract). J Trauma 1999;47:221.
155. Mullins RJ, Mann NC, Worrall W, Jurkovich GJ, and Hedges JR. Preferential benefit from implementation of a statewide trauma system in one of two adjacent states (abstract). J Trauma 1997;43:122.
156. Nicholl J and Turner J. Effectiveness of a regional trauma system in reducing mortality from major trauma: before and after study. BMJ 1997;315:1349-1354.
157. Stewart TC, Lane PL, and Stefanits T. An evaluation of patient outcomes before and after trauma center designation using trauma and injury severity score analysis. J Trauma 1995;39:1036-1040.
158. Karsteadt LL, Larsen, CL, and Farmer PD. Analysis of a rural trauma program using the TRISS methodology: A three-year retrospective study. J Trauma 1994;36:395-400.
159. Hill DA, West RH, and Duflou J. Value of the prospective 'before and after' study as a methodology to evaluate outcome in a trauma centre. Aust NZ J Surg 1993;63:940-945.
160. Champion HR, Sacco WJ, and Copes WS. Improvement in outcome from trauma center care. Arch Surg 1992;127:333-338.
161. Shackford SR, Mackersie RC, Hoyt DB, Baxt WG, Eastman AB, Hammill FN, Knotts B, and Virgilio RW. Impact of a trauma system on outcome of severely injured patients. Arch Surg 1987;122:523-527.
162. Cales RH. Trauma mortality in Orange County: the effect of implementation of a regional trauma system. Ann Emerg Med 1984;13:1- 10.
163. West JG, Cales RH, and Gazzaniga AB. Impact of regionalization. The Orange County experience. Arch Surg 1983;118:740-744.
164. West JG, Trunkey DD, and Lim RC. Systems of trauma care. A study of two counties. Arch Surg 1979;114:455-460.
165. Waters JM and Wells CH. The effects of a modern emergency medical care system in reducing automoblile crash deaths. J Trauma 1973;13:645-647.29.

Moishe Liberman MD, is a fourth year resident in the Division of General Surgery at McGill University. He is also a research fellow in the Department of Experimental Surgery at McGill. **David S. Mulder MD, MSc, FRCSC** is the Chairman of the Division of Cardiothoracic Surgery at the McGill University Health Center, Senior Surgeon at the Montreal General Hospital, and Professor of Surgery at McGill University. **John S. Sampalis PhD,** is the Head of Surgical Epidemiology at McGill University, an Associate Professor in the Departments of Surgery, Epidemiology and Biostatistics, and an Associate Professor in the Department of Social and Preventive Medicine at Université de Montréal and at Université de Laval. He has been a past recipient of the Medical Research of Canada Scientist award and is currently a senior scientist supported by the Fonds de la recherche en sante du Quebec (FRSQ).

MJM

AN INTERNATIONAL FORUM FOR THE ADVANCEMENT OF MEDICAL SCIENCE BY STUDENTS

INSTRUCTION FOR AUTHORS

The *McGill Journal of Medicine* (*MJM*) provides an international forum for student contributions to the field of medicine. The scope of *MJM* articles encompasses clinical medicine (clinical trials, case reports, surgery, epidemiology, etc.), basic science (biochemistry, cell biology, physiology, cancer research, etc.), and the humanities (ethics, health policy, history of medicine, etc.). Submissions of original research articles, review articles, and articles relating medicine to the humanities ("Crossroads" articles) are encouraged. Submissions of single clinical case reports and special features are also welcome, as are letters and commentaries on previously published articles, discussing various medical issues, or communicating preliminary research results (see below).

AUTHORSHIP

The *MJM* publishes original, medically-relevant articles not published or currently submitted elsewhere. As the *MJM* is a journal of student research, the student author must have participated sufficiently in the work to take public responsibility for its content, as well as to appear as first author. Editors may require contributors to justify the assignment of authorship.

MANUSCRIPT STYLE

Manuscripts must be in the English language. The *MJM* conforms to the standards for style defined in the CBE Style Manual, 5th ed., Council of Biology Editors, Inc., Bethesda, MD, 1983.

TYPES OF ARTICLES

Non-peer reviewed

Letter to the MJM – short comment on a specific topic, on a paper previously published in the *MJM*, or on the *MJM* itself. Unsolicited letters are welcome.

Commentary – longer than letters but generally consisting of fewer than 700 words, on a topic of general interest. Unsolicited Commentaries are welcome but authors should contact the Editor-in-Chief to discuss the topic before submitting.

Focus Article – solicited articles on a subject chosen in advance by the Editorial Board.

Book Review – comments on a book of general interest. Recently published books are preferred but older books still considered to have wide appeal are appropriate. Unsolicited book reviews are welcome but potential authors should contact the Editor-in-Chief to discuss the book before submitting.

Peer reviewed (First author must be a student)

Original Article – full-length article reporting the results of an original research project. In addition to presenting the project and its findings, the project's rationale and the significance of its findings must be clearly explained.

Case Report – report and discussion of a clinical case that illustrates some important teaching points. Authors must clearly explain the significance of the chosen case and the teaching points illustrated by it.

Review Article – overview of the understanding and outstanding questions in a particular field of research. Review articles that explore a specific hypothesis and/or combine information from different areas of research to advance an original idea are especially encouraged, as opposed to review articles that simply enumerate past findings.

Crossroads – similar in style to review articles. "Crossroads" articles should explore the relationship between medicine and the humanities (visual arts, literature, history, philosophy, etc.).

ORGANIZATION

For all peer reviewed articles, each required component should begin on a new page and appear sequentially, as follows. Original articles: title page, abstract, introduction, methods, results, discussion, acknowledgements, references, tables, and illustrations. Case report: title page, introduction, the case, discussion, acknowledgments, references, tables, and illustrations. Review and "Crossroads" articles: title page, introduction, body of text, conclusion, acknowledgments, references, tables, and illustrations.

Tables

Tables must be submitted typewritten in the order corresponding to their first citation in the text and accompanied by brief titles. All non-standard abbreviations must appear below the table with an accompanying definition.

Illustrations

Figures and legends should contain sufficient information such that each figure is intelligible without reference to the text. Figures must be printed on high-quality paper. Photographs should be glossy in black and white. Authors may submit color prints but, if published, must bear the extra printing costs. Each figure legend (with figure title and corresponding figure number) should be submitted on a separate page.

REFERENCES

Reference citations should appear in numerical order in parentheses throughout the text and be listed in their order of appearance. References should be formatted according to the examples below.

Journal articles

1. Bunny B, Coyote WE, Le Pew P, et al. Impact trauma Caused by Descending GPI Anchors. McGill Journal of Medicine 2003; 7: 1-2.

Books

2. Bunny B, Coyote WE, Le Pew P. Subdural Hematomas. In: Jones J, ed. Head Injuries. New York: Acme Publishers; 1994: 249-260.

Internet

3. Bunny B. Computer-induced psychosis. Society for Cartoon-Computer Interactions. http://www.SCarComI.com/psychosis.html. 1999.

EDITORIAL REVIEW PROCESS

All submissions of "peer reviewed articles" undergo a formal and anonymous evaluation process conducted by the *MJM* Editorial Board, comprised of trained students with diverse areas of expertise. This peer review process is performed in two stages. During the first stage, the manuscript is evaluated individually by each member of the Editorial Board. For submissions requiring revision, a detailed manuscript evaluation summary is returned to the author with suggestions for improvement. Manuscripts of high caliber proceed to the second stage, in which a review is performed by a McGill medical faculty member whose field of expertise correlates with the subject matter of the paper. Comments from the faculty expert are synthesized with the evaluations of the Editor-in-Chief, Executive and Senior editors, and members of the Editorial Board to achieve a decision to accept, defer, or reject the paper for publication. In all cases, a detailed manuscript evaluation summary is returned to the author along with any suggestions for revision.

After appropriate revision, the manuscript text must be resubmitted on diskette and a copyright transfer agreement signed. Prior to publication, authors will be required to review edited proofs without delay.

SUBMISSION OF MANUSCRIPTS

Manuscripts must be submitted by e-mail in MS Word format to *mjm.med@mcgill.ca*. You may contact the McGill Journal of Medicine through the Faculty of Medicine, McGill University, 6th floor 3655 Promenade Sir William Osler, Montreal, Quebec, Canada H3G 1Y6. The *MJM* Editorial Division may also be contacted by telephone at (514) 398-6987, or by e-mail at *mjm.med@mcgill.ca*.

This is an abridged version of the instructions to authors. For the full version, visit our Website, http://www.mjm.mcgill.ca or contact us as indicated above.

MJM

AN INTERNATIONAL FORUM FOR THE ADVANCEMENT OF MEDICAL SCIENCE BY STUDENTS

THERE IS ONLY ONE.

Read the work of current and future medical scientists.
Join the *MJM* today.

Please photocopy and send in the order form below.

YES! Please begin my one year subscription to the *MJM**.

Name _____

Address _____

City _____

Province/State _____

Postal/Zip Code _____ Country _____

Please start subscription with: Volume ____ Issue ____

		STUDENT	NON-STUDENT
*	Prices for one year subscription (2 issues)	CAN$20 (Canada) US$15 (USA) US$20 (international)	CAN$40 (Canada) US$30 (USA) US$40 (international)

All rates include postage, handling, and 7% GST. Please make checks payable to: *MJM*.
Mail to: Subscription Requests, *MJM*, Faculty of Medicine, McGill University, 3655 Promenade Sir William Osler, Montreal, QC, Canada H3G 1Y6.