

Global strategies to reduce the health-care burden of craniofacial anomalies

*Report of WHO meetings on International
Collaborative Research on Craniofacial Anomalies*

Geneva, Switzerland, 5-8 November 2000

Park City, Utah, USA, 24-26 May 2001



Human Genetics Programme – 2002
Management of Noncommunicable Diseases
World Health Organization
Geneva, Switzerland

THIS REPORT IS DEDICATED TO THE MEMORY OF
DR DAVID BARMES,
A FORMER STAFF MEMBER OF THE WORLD HEALTH ORGANIZATION
AND FOUNDER OF THE INTERNATIONAL COLLABORATIVE RESEARCH
PROJECT ON CRANIOFACIAL ANOMALIES.

Acknowledgements: These meetings were organized by the World Health Organization, with financial support from the National Institute of Dental and Craniofacial Research (NIDCR) of the National Institutes of Health (United States). In particular, the participants of the meetings wish to express their sincere thanks to Dr Kevin Hardwick and Dr Rochelle Small from NIDCR for their support and advice.



WHO Library Cataloguing-in-Publication Data

Global strategies to reduce the health-care burden of craniofacial anomalies : report of WHO meetings on International Collaborative Research on Craniofacial Anomalies, Geneva, Switzerland, 5-8 November 2000 ; Park City, Utah, U. S. A., 24-26 May 2001.

1.Craniofacial abnormalities – etiology 2.Craniofacial abnormalities – therapy 3.Cleft lip – epidemiology 4.Cleft palate – epidemiology 5.Evidence-based medicine – standards 6.Cost of illness 7.Research 8.Practice guidelines

I.WHO Meeting on International Collaborative Research on Craniofacial Anomalies (1st : 2000 : Geneva, Switzerland)

II.WHO Meeting on International Collaborative Research on Craniofacial Anomalies (2nd : 2001 : Park City, Utah)

ISBN 92 4 159038 6 (NLM classification:WE 705)

© 2002 World Health Organization

All rights reserved. Publications of the World Health Organization can be obtained from Marketing and Dissemination, World Health Organization, 20 Avenue Appia, 1211 Geneva 27, Switzerland (tel: +41 22 791 2476; fax: +41 22 791 4857; email: bookorders@who.int). Requests for permission to reproduce or translate WHO publications – whether for sale or for noncommercial distribution – should be addressed to **Publications**, at the above address (fax: +41 22 791 4806; email: permissions@who.int).

The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the World Health Organization concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by the World Health Organization in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters.

The World Health Organization does not warrant that the information contained in this publication is complete and correct and shall not be liable for any damages incurred as a result of its use.

This report contains the collective views of an international group of experts, and does not necessarily represent the decisions or the stated policy of the World Health Organization.

Cover by WHO Graphics.

Printed in France.

Contents

<i>List of boxes</i>	<i>v</i>
<i>Acronyms and abbreviations</i>	<i>vii</i>
<i>Executive summary</i>	<i>ix</i>
1. Introduction	1
2. Global epidemiology and health burden of CFA	4
2.1 Global epidemiology	4
2.2 Recommendations for producing better descriptive statistics in OFC	11
2.3 Conclusions	13
3. Possibilities for improving the treatment of CFA	14
3.1 Evidence-based care	14
3.2 Quality improvement	32
3.3 Access and availability	36
3.4 Regional perspectives	38
4. Role of environment in CFA	41
4.1 Socioeconomic status and orofacial clefts	41
4.2 Nutrition and orofacial clefts: general issues	42
4.3 Folic acid: nutritional biochemistry and orofacial clefts	45
4.4 Other specific nutrients and orofacial clefts	49
4.5 Lifestyle, occupational and other environmental factors in orofacial clefting	50
4.6 Conclusions	53
5. Role of genes in CFA	54
5.1 Embryogenesis	55
5.2 Clinical definition of craniofacial anomalies	57
5.3 Mouse models	59
5.4 Genotyping	60
5.5 Recent developments	66

6.	Gene/environment interactions	67
6.1	Introduction	67
6.2	Data challenges	68
6.3	Methodology challenges	70
6.4	Collaboration challenges	71
6.5	Conclusions	73
7.	Prevention of CFA	74
7.1	Meeting objectives	74
7.2	Environmental and behavioural factors and orofacial clefts	75
7.3	Maternal nutrition and orofacial clefts	77
7.4	Nutritional supplementation	80
7.5	Ethical issues	84
7.6	The design of orofacial-cleft prevention trials	92
7.7	Conclusions	94
8.	Conclusions and recommendations	96
8.1	Treatment of CFA	96
8.2	Gene/environment interaction	98
8.3	Genetics	100
8.4	Prevention	102
9.	List of participants	104
10.	References	114
	Annex 1: European Collaboration on Craniofacial Anomalies (EUROCRAN)	137
	Annex 2: Work packages	138
	Annex 3: Policy statements	140
	Annex 4: Practice guidelines	142
	Annex 5: General principles governing record-taking (provisional)	144

List of boxes

Box A	Recommendations for producing better descriptive statistics in OFC and epidemiology	12
Box B	Sources of bias in CFA research	17
Box C	Systematic planning for surgical trials	27
Box D	Measuring treatment burden	28
Box E	Priorities for research on treatment	31
Box F	International recommendations on organization of cleft lip and palate services	33
Box G	Worldwide variation in diet	43
Box H	Maternal cigarette smoking in pregnancy	50
Box I	Strategic approaches	60
Box J	Collaborative strategies	65
Box K	Gaps and challenges in the study of gene/environment interactions in orofacial clefting	68
Box L	More people, more countries	72
Box M	Objectives of the WHO meeting on the prevention of CFA	74
Box N	Trials of maternal nutritional supplementation and orofacial clefts	80
Box O	Future directions for orofacial-cleft prevention trials	83
Box P	Ethical issues related to studies of maternal nutrition and birth defects	84
Box Q	Equipoise	86
Box R	Highest attainable and sustainable standard of care	91
Box S	Role for WHO, governmental agencies and non-governmental organizations	95

Acronyms and abbreviations

AED	anti-epileptic drugs
AZT	azidothymidine
BCLP	bilateral cleft lip and palate
CAPS	Cleft Audit Protocol for Speech
CAT (scan)	computerized axial tomography
CDC	Centers for Disease Control and Prevention (United States of America)
CFA	craniofacial anomalies
CI	confidence interval
CIDR	Centre for Inherited Disease Research
CIOMS	Council for International Organizations of Medical Science
CL	cleft lip
CL/P	cleft lip – with or without cleft palate
CLP	cleft lip and palate
COR	Craniofacial Outcomes Registry
CP	isolated cleft palate
CPAP	continuous airway pressure
DNA	deoxyribonucleic acid
ECLAMC	Estudio Colaborativo Latino Americano Malformaciones Congenita
ENT	ear, nose and throat
ESF	European Science Foundation
EU	European Union
EUROCAT	European Registry for Congenital Anomalies and Twins
EUROCRAN	European Collaboration on Craniofacial Anomalies
FAS	fetal alcohol syndrome
FFQ	food-frequency questionnaire
GEI	gene/environment interaction

GOS.SP.ASS	Great Ormond Street speech assessment
HIV	human immunodeficiency virus
IARC	International Agency for Research on Cancer
IFR6	interferon regulatory factor 6
IMR	infant mortality rate
IU	international units
LRT	likelihood ratio test
mg	milligrams
MRC	Medical Research Council (United Kingdom)
MRI	magnetic resonance imaging
MSX	muscle-specific homeobox factor
NGO	non-governmental organization
NIH	National Institutes of Health (United States of America)
NTD	neural tube defects
OFC	orofacial clefts
OR	odds ratio
RCT	randomized controlled trial
RNA	ribonucleic acid
RR	relative risk
TCS	Treacher Collins syndrome
TDT	transmission disequilibrium test
TGF	transforming growth factor
UCLP	unilateral cleft of the lip and palate
VCF	velo-cardio-facial syndrome
VPI	velo-pharyngeal incompetence
WHO	World Health Organization
WMA	World Medical Association

Executive summary

In 2000, the WHO Human Genetics Programme, with financial support from the United States National Institute of Dental and Craniofacial Research, launched a five-year project designed to take forward an international research strategy on craniofacial anomalies (CFA). The specific objectives of this initiative are:

- to develop an international network for consensus building, planning and protocol development for international, collaborative, biomedical, epidemiological and behavioural studies in the core areas of CFA research;
- to create a directory of CFA research resources, and
- to establish a publicly-accessible research database on the Internet.

As a first step of this initiative, a consensus conference of international experts covering the four selected areas for research – treatment of CFA, gene/environment interaction (GEI), genetics, and prevention – was held under the auspices of the World Health Organization (WHO). The conference comprised two meetings – the first, held in Geneva from 5-8 November 2000, included concurrent workshops on research concerning the genetic basis of CFA, gene/environment interactions, and the treatment of CFA; the second, held in Utah from 24-26 May 2001, considered the prevention of CFA.

The aims and objectives of the WHO consensus meetings were to:

- (1) obtain counsel from experts involved in CFA research around the world;
- (2) describe the “state-of-the-science” with regard to treatment, genetics, gene/environment interaction and prevention, and highlight recent relevant research;
- (3) discuss requirements for future research in all areas of craniofacial anomalies; and

-
- (4) arrive at a consensus on approaches to address data gaps and proceed with strategies, methodologies and protocols to advance knowledge.

A. Treatment

Three interrelated research issues were addressed within the clinical theme:

- (1) **Evidence-based care:** the identification and dissemination of optimal clinical interventions for the management of CFA.
- (2) **Quality improvement:** the development and dissemination of methodologies for monitoring and improving the delivery of clinical services.
- (3) **Access and availability:** the identification of strategies to maximize access to adequate levels of care for all affected individuals, irrespective of nationality.

B. Gene/environment interaction

Issues discussed in relation to the planning of future collaborative gene/environment interaction (GEI) research were:

■ Identification of data gaps

- (1) Use birth surveillance systems to determine the frequency of craniofacial anomalies and sources in ascertainment.
- (2) Identify areas of the world where interesting populations or patterns of craniofacial anomalies exist, and gain access to those populations.
- (3) Evaluate whether an established infrastructure exists to allow research in GEI to proceed.
- (4) For GEI research it will be essential to carefully categorize samples by type of defect, to identify (and exclude) syndromes that are known to have a genetic etiology and, where possible, to control methodologic and demographic parameters which might confound biochemical and genetic analyses. This type of research is therefore predominantly applied to non-syndromic orofacial clefts.
- (5) GEI research should seek to establish the frequency of genotypes in different populations and ethnic groups and establish the risk of orofacial clefts associated with:
 - (a) the gene variant alone,
 - (b) environmental exposures alone, and
 - (c) gene/environment interaction.

■ Study design and standardization issues

Having identified data gaps, appropriate research hypotheses can be generated. Agreement will be required on the data to be collected, the methods of sample collection and the geographical areas where research would be carried out. In time it would be anticipated that the research would address the data gaps identified and would raise further issues that would be addressed by generating further hypotheses to be tested in a cycle of enquiry and research.

■ Common core protocols

It was agreed that the standardization of research would require the development of guidelines to provide consistency between groups collecting data. Such common core protocols would be developed in the areas of:

- (a) nutritional, lifestyle and occupational factors;
- (b) medical, obstetric and drug histories;
- (c) genetic and biochemical data collection;
- (d) assessment of clinical dysmorphology and collection of consistent family history data;
- (e) agreed guidelines for ascertainment of cases and, where appropriate, controls.

C. Genetics

While there is an inevitable overlap between research in genetics and in gene/environment interaction, CFA research will benefit from an intensive genetics approach.

- (1) The discussions on the genetics component of the WHO CFA Conference focused on those technologies, analytic approaches, and populations that will best advance our understanding of the etiologies of craniofacial abnormalities, with particular reference to those with strong genetic components.
- (2) While recognizing that the environment and stochastic events play an important and, often, major role in predisposing to craniofacial anomalies, the role of genetics is compelling in many situations.
- (3) Funding, manpower training, bioethical and government policy issues also influence research. These should be discussed and addressed in the light of identified differences in the demographics and infrastructure in different regions, and research priorities should be established geographically and according to agreed criteria.

D. Prevention

- (1) Identify environmental and behavioural factors with established associations with orofacial clefts and other CFA.
- (2) Review evidence on the role of specific maternal nutritional factors in the etiology of orofacial clefts and other CFA.
- (3) Reach a consensus regarding the role and importance of nutritional supplementation trials in evaluating the causal role of specific nutrients in the etiology of orofacial clefts and other CFA.
- (4) Discuss aspects of the design of orofacial cleft and CFA prevention trials and their ethical, legal, social and financial implications.
- (5) Make recommendations on the resources needed to implement international collaborative studies of CFA prevention with common core protocols.

Section 8 provides details of the recommendations for future research arising out of these two WHO consensus meetings.

1

Introduction

Craniofacial anomalies (CFA) are a highly diverse group of complex congenital anomalies. Collectively they affect a significant proportion of the global society (*see Table 1 below*).

Table 1: Examples of most common craniofacial anomalies

	Prevalence at birth: per 10 000
Cleft lip ± palate	
Caucasian	10
Japanese	20
Native (North) Americans	36
African American population	3
Cleft palate	
Averaged across races	5
Craniosynostosis	
Crouzon syndrome	0.4
Apert syndrome	0.15
Otomandibular anomalies	
Treacher Collins syndrome	0.2
CHARGE Association	
	1
Holoprosencephaly	
	1.2
Stickler syndrome	
	1
Fetal alcohol syndrome	
	2

Source: Rovin et al., 1964; Temple, 1989; Cohen et al., 1992; Lewanda et al., 1992; Croen et al., 1996; Derijcke et al., 1996; Sampson et al., 1997; Blake et al., 1998.

Great confusion surrounds the optimal management for even the most common conditions

The prevalence of individual conditions varies considerably across geographic areas and ethnic groupings. Their impact on speech, hearing, appearance and cognition has a prolonged and adverse influence on health and social integration. The costs incurred from CFA in terms of morbidity, health care, emotional disturbance, and social and employment exclusion are considerable for affected individuals, their families and society. Research that will increase the understanding of the causes of CFA, improve the treatment for it, and lead ultimately to its prevention or reduction, has mainly been pursued in the absence of an international strategy. Yet international collaboration is a prerequisite for accessing adequate samples for research in etiology, treatment and prevention, and also for the assembly of a critical mass of clinical researchers and basic scientists in fields such as molecular biology, genetics, biochemistry and epidemiology.

The treatment of CFA has, so far, escaped the rigours of contemporary health technology assessment, and great confusion surrounds the optimal management for even the most common conditions. For each of the many subgroups of CFA, the attainment of homogeneous samples of adequate size for randomized trials and long-term follow-up represents a formidable challenge. Multi-site cooperation is essential. In the developing world, the costs of rehabilitation and problems of access put treatment beyond the reach of vast numbers of affected individuals. Systems for delivering care in different geographic and economic circumstances urgently require research.

The potential of research on the genetic basis of CFA has increased dramatically over the last decade with the development of recombinant DNA technology. In over 50 craniofacial syndromes, genes involved have either been mapped to a chromosome location or actively isolated and their structure identified. This achievement, however, represents only a fraction of the total number of craniofacial syndromes defined. The pathogenesis of the most common forms of CFA – non-syndromic clefts of lip and/or palate – is especially challenging because they appear to arise from complex polygenic interactions with environmental factors. A coordinated international approach would not only provide effective means of sharing data, samples and resources, but would allow strategic exploitation of geographic and ethnic variation in the incidence and pathogenesis of CFA.

Research that may lead to the prevention of CFA has been based, primarily, on isolated case control studies in Asia, Europe, Latin America and the United States of America. As yet, these projects have occurred independently of each other, and consistent conclusions about viable interventions such as dietary supplementation in the periconceptual

period have yet to emerge. Once again, international standardization of research protocols, consensus on preventive interventions suitable for clinical trials, and the performance of trials in an international framework, would enhance the validity, consistency and generalizability of these efforts.

Efforts to define an international research strategy go back more than a decade when the proposals for “International Collaboration on Oral Health” were jointly published by WHO, the International Dental Federation (FDI), and the US National Institute for Dental and Craniofacial Research. More recently these proposals were renewed at a series of consensus meetings:

- Eighth Congress of the International Confederation of Craniofacial Teams, Singapore, 1997;
- Craniofacial Genetic Diseases and Disorders Planning Workshop, Bethesda, USA, 1997;
- International Collaboration on Oral Cleft Genetics Second Meeting, Baltimore, USA, 1998; and
- Meeting of the International Task Force on CFA, Bauru, Brazil, 1998.

In 2000, the WHO Human Genetics Programme, with financial support from the US National Institute of Dental and Craniofacial Research, launched a five-year project designed to take these proposals forward. The specific objectives of this initiative have been to develop an international network for consensus building, planning and protocol development for international, collaborative, biomedical, epidemiological and behavioural studies in the core areas of CFA research, and to create a directory of CFA research resources and a publicly-accessible research database on the Internet.

This report is based on the first two consensus meetings of international of experts held under the auspices of WHO. The first meeting, held in Geneva, 5-8 November 2000, included concurrent workshops on research concerning the genetic basis of CFA, gene/environment interactions, and the treatment of CFA. The second meeting, held in Utah, 24-26 May 2001, considered the prevention of CFA.

2

Global epidemiology and health burden of CFA

2.1 Global epidemiology

Cleft lip, with or without cleft palate (CL/P), and isolated cleft palate (CP) are serious birth defects which affect approximately 1 in every 600 newborn babies worldwide. This means that, assuming 15 000 children are born per hour worldwide (United States Bureau of the Census, 2001), a child is born with a cleft somewhere in the world approximately every 2½ minutes. From birth to maturity, children with orofacial clefts (OFC) undergo multidisciplinary surgical and non-surgical treatment with considerable disruption to their lives, and often with adverse psychological consequences to themselves and their families.

Over the years efforts have been made to record frequency of birth defects. Accurate data on the epidemiology are important not only for documenting the burden in relation to the planning of public health services, but also because they form the basis for research into the causes. The eventual objective, from both scientific and humanitarian viewpoints, must be to advance the knowledge and understanding of causative factors so as to be able to institute primary preventive measures. Among the barriers to achieving this objective are: (a) the heterogeneity of orofacial clefting; (b) the lack of standard criteria for the collection of data; and (c) in particular the lack of and/or failure to apply an internationally comparable classification for orofacial clefting.

The level of ascertainment differs between countries, depending on the method of cleft birth registration; the number of live births, terminations, stillbirths and syndromic individuals can considerably affect the validity of such data. The critical requirement is to precisely define the "population" in which malformations are measured. The main issue is whether one reports or estimates rates in all conceptuses, all births, or all live births. The word *births* is somewhat ambiguous because it usually includes stillbirths, a term which does not have a uniform definition.

2.1.1 Epidemiological data summary

Epidemiological data for orofacial clefts from the three different sources outlined above are presented in peer-reviewed publications. Tables 2 and 3 (WHO, 1998) show data from the peer-reviewed literature and that collected through the International Clearinghouse Birth Defects Monitoring System (ICBDMS) and European Registration of Congenital Anomalies (EUROCAT).

Birth prevalence studies on patients with CL/P and CP over the second half of the 20th century reveal that whilst there are ethnic and geographic differences, the "average" birth prevalence of orofacial clefting in the world's western populations is often quoted as 1:1000 total births for CL/P and 1:2000 total births for CP (*see Tables 2 and 3*). The birth prevalence of CL/P is highest in Australia (Aborigines), Canada, the Far East, India, Scandinavia, parts of South America, and the USA, and lower in Southern Europe. In general populations of Asian origin have a higher incidence than Caucasian populations which, in turn, have a higher incidence than African populations. The birth prevalence of CL/P varies from 2.7:1000 in Native Americans to 2.1:1000 in Japan and to 0.4:1000 in Nigeria and 0.42:1000 in African Americans (Leck, 1972), with the geographical variation being less important than ethnic differences.

Cleft palate alone (CP) has a lower average birth prevalence and shows less variation in different racial groups. The prevalence of CP is highest in Australia, Finland, and Scotland (United Kingdom), and in general is higher in Asians than Caucasians or Africans (Melnick, 1992). Generally CL/P occurs more frequently in males whereas for CP the reverse is true. Significant racial differences in the birth prevalence of orofacial clefts exist. Two thirds of all cases of unilateral CL/P have left-sided defects regardless of gender, race and severity of defect (Fraser and Calnan, 1961).

Migrants studies show that African Americans have lower rates for both CP and CLP than Whites in the United States, and a study in Birmingham (United Kingdom) also showed that those originating from the Caribbean have low rates of orofacial clefting (Leck, 1969; Leck and Lancashire, 1995). Studies in North America also reveal similar rates among Japanese-Americans and Chinese-Americans compared to Caucasian-Americans (Croen et al, 1998); there is also evidence that the frequency of CL/P (but not CP) may be significantly lower among US-born Japanese and other Asians born in California and New York than among those born in Japan or Hawaii (Tyan, 1982). The worldwide variation in the frequency of orofacial clefts (OFC) is likely therefore to be influenced by the variable predisposing factors that exist, depending on ethnicity and geography. When comparing the data, however, it is important to consider issues which affect the figures, such as: (a) statistical variability of recorded rates; (b) live births versus stillbirths; and (c) associated malformations.

Table 2: Cleft lip with or without cleft palate

	Live and stillbirths	Induced abortions	Total cases	Total births	Rates (per 10 000)	
Argentina	99	– (*)	99	73 942	13.4	↑
Australia – South Australia	–	–	19	19 801	9.6	
Australia – Victoria	26	47	73	65 182	11.2	
Belarus	–	–	–	–	–	
Belgium – Hainaut Namur	30	1	31	24 856	12.5	
Brazil	51	– (*)	51	36 689	13.9	
Chile	20	– (*)	20	22 276	9.0	
Czech Republic	113	–	113	107 153	10.5	
Denmark – Odense	17	0	17	12 054	14.1	
France – Bouches du Rhone	33	3	36	44 704	8.1	
France – Central East	74	4	78	100 074	7.8	↓
France – Paris	47	16	63	71 319	8.8	
France – Strasbourg	29	5	34	27 200	12.5	
Ireland – Dublin	31	– (*)	31	38 000	8.2	
Italy – Campania	38	2	40	43 325	9.2	
Italy – Emilia Romagna	25	–	25	25 924	9.6	
Italy – Toscana	42	1	43	48 991	8.8	
Japan	172	– (*)	172	113 702	15.1	↑
Mexico	81	– (*)	81	65 870	12.3	
Netherlands – North	52	6	58	38 670	15.0	↑
Norway	99	2	101	60 584	16.7	↑
Spain – Basque Country	114	2	16	31 248	5.1	↓
Switzerland	101	4	105	148 000	7.1	↓
United Kingdom – Belfast	10	1	11	49 482	2.2	↓
United Kingdom – Glasgow	19	1	20	22 570	8.9	
United Kingdom – North Thames	43	10	53	47 274	11.2	
USA – Atlanta	34	0	34	39 856	8.5	
USA – Hawaii	–	–	22	20 596	10.7	
Uruguay	17	– (*)	17	21 332	8.0	
Venezuela	21	– (*)	21	36 377	5.8	↓

- * Abortion for birth defect not permitted.
 ↑ = 99% significantly higher than the mean.
 ↓ = 99% significantly lower than the mean.

Source: WHO (1998) World Atlas of Birth Defects (1st edition)

Table 3: Cleft palate without cleft lip

	Live and stillbirths	Induced abortions	Total cases	Total births	Rates (per 10 000)	
Argentina	43	– (*)	43	73 942	5.8	
Australia – South Australia	18	–	18	19 801	9.1	
Australia – Victoria	39	0	39	65 182	6.0	
Belarus	–	–	–	–	–	
Belgium – Hainaut Namur	15	2	17	24 856	6.8	
Brazil	19	– (*)	19	366 689	5.2	
Chile	13	– (*)	13	22 276	5.8	
Czech Republic	66	–	66	107 153	6.2	
Denmark – Odense	11	0	11	12 054	9.1	
France – Bouches du Rhone	23	5	28	44 704	6.3	
France – Central East	72	7	79	100 074	7.9	↑
France – Paris	36	14	50	71 319	7.0	
France – Strasbourg	21	2	23	27 200	8.5	
Ireland – Dublin	13	– (*)	13	38 000	3.4	
Italy – Campania	24	–	24	43 325	5.5	
Italy – Emilia Romagna	12	–	12	25 924	4.6	
Italy – Toscana	10	2	12	48 991	2.4	↓
Japan	52	– (*)	52	113 702	4.6	
Mexico	27	– (*)	27	65 870	4.1	
Netherlands – North	32	1	33	38 670	8.5	
Norway	26	0	26	60 584	4.3	
Spain – Basque Country	17	1	18	31 248	5.8	
Switzerland	63	3	66	148 000	4.5	
United Kingdom – Belfast	6	1	7	49 482	1.4	↓
United Kingdom – Glasgow	19	3	22	22 570	9.7	
United Kingdom – North Thames	20	2	22	47 274	4.7	
USA – Atlanta	12	1	13	39 856	3.3	
USA – Hawaii	12	–	12	20 596	5.8	
Uruguay	10	– (*)	10	21 332	4.7	
Venezuela	14	– (*)	14	36 377	3.8	
Total			789	1 457 051	5.4	

Source: WHO (1998) World Atlas of Birth Defects (1st edition)

* Abortion for birth defect not permitted.

↑ = 99% significantly higher than the mean.

↓ = 99% significantly lower than the mean.

2.1.2. *Variability of recorded rates*

The precision of recorded rates depends on the recording of the total population birth rate (denominator data) and the recognition and recording of the number of affected births. Since the incidence and birth prevalence of OFC is low, the variability of the rate depends primarily on the level of ascertainment and number of abnormal births recorded. The standard error of the observed number x (Poisson distribution) is simply its square root (\sqrt{x}) and the width of the 95% confidence limit for x is $1.96 \sqrt{x}$. The width of the confidence interval as a percentage of the observed number is a measure of the precision. Studies that have a statistical variability of more than 30%, however, need to be interpreted with caution.

There is
apparent variation
in the proportion
of OFC cases
with additional
congenital
anomalies and
syndromes

Many of the studies described in developing countries are based on hospital rather than general population figures so will only be accurate in communities where it is likely that the vast majority of births have occurred in hospital. In the interests of recording reasonably accurate data, information from registries only is displayed above, and the figures for some studies in Africa, India and the Middle East are excluded.

2.1.3. *Live births versus stillbirths*

The proportion of serious malformations is higher in stillbirths than in live births so including stillbirths tends to raise the birth prevalence or incidence rates above those that only consider live births. Similarly, inclusion of data on earlier loss – miscarriages and abortions – will increase rates over data that analyse only live births and stillbirths.

Vanderas (1987) examined the problem of inclusion or exclusion of stillbirths as an issue in ascertainment of OFC in a number of international studies, some of which included live births, stillbirths and abortions in their evaluation of incidence rate. The OFC rates were 6.43 per 1000 stillbirths versus 2.16 per 1000 live births in Hay's study (1971) of Caucasians in the United States (Iowa); and 2.72 per 1000 stillbirths versus 0.91 per 1000 live births in the pooled data Lutz and Moore (Lutz et al., 1955) compiled on African Americans, Mexicans and Caucasians. It appears, therefore, that in stillbirths and abortions the risk of developing clefts is about three times more frequent than in live births; and clefts with associated malformations behave differently epidemiologically from clefts without associated malformations.

A further study in Hungary (Czeizel et al., 1984) reported that the proportion of cleft palate without cleft lip is about sevenfold greater in stillbirths (primary fetal deaths 28 weeks or older) than in live births (2.38 per 1000 versus 0.36 per 1000). Whereas for cleft lip (with or without cleft palate), the ratio is a little less than threefold (3.17 per 1000 versus

1.15 per 1000). As may be expected, this differential between live births and stillbirths is greater for those orofacial clefts that occur in individuals with additional malformations elsewhere, than in those with only cleft lip, cleft palate, or both.

Krause (1963) examined human embryos and fetuses and reported that the frequency of clefts with associated malformations was 11.61 per 1000, and fetuses with clefts but without associated malformations were 7.22 per 1000. Nishimura (1966), reported the frequency of cleft lip with or without cleft palate in 1213 voluntarily aborted human embryos in Japan to be 14.7 per 1000. In a later Japanese study on 5117 voluntarily aborted human embryos, Iizuka (1973), found that the incidence of cleft lip (CL) was 4.3 per 1000, cleft lip and palate (CLP) 8.1 per 1000 and isolated cleft palate (CP) 3.2 per 1000.

It is for this reason that the indiscriminate grouping of figures which include not only live births but also stillbirths and/or induced abortions will not be comparable to those which quote live births only. If fetal deaths or earlier losses are included in summary rates, this should be noted specifically and rates should be presented separately for live births and for embryonic and fetal deaths.

2.1.4 Associated malformations

It is generally accepted that associated malformations occur more frequently in infants who have CP than in those who have CLP and even less still in those with isolated CL. For example, a 17-year study in North Eastern France reported the rate of associated malformations as 46.7% in CP, 36.8% in CLP and 13.6% in CL (Kallen et al., 1996). Cornel (1992) reported associated abnormalities in 23% of combined CL/P cases and in 52% of cases with isolated CP. Other studies that also found congenital anomalies to be much more commonly associated with CP than with CL/P were Ingalls et al., 1964; Drillien et al., 1966; Moller, 1972 and Emanuel et al., 1973. In the Finnish population, however, CL/P was as often associated with other malformations as was CP (Saxen et al., 1974). Familial background was also more often reported in association with CP than with CL/P in Finland; this is in contrast to that found by others, such as Fogh-Andersen (1942) in Denmark.

Some reports also sub-divide CL/P into unilateral and bilateral sub-groups when examining additional malformations and report an increase in additional malformations in the bilateral sub-group (e.g. Hagberg et al., 1997). When considering associated abnormalities some reports do not define what is meant by "associated abnormalities" while others give ambiguous descriptions, and Conway and Wagner (1966) record only the "10 most common" associated abnormalities listed on birth certificates over an 11-year period.

2.1.5 *The prevalence of isolated cleft palate*

There is considerable international variation in the frequency of OFCs

There is considerable heterogeneity in what is described as isolated cleft palate. Many figures for isolated cleft palate are provided without an adequate explanation of inclusion/exclusion criteria. For instance, the most common syndrome with isolated cleft palate as a feature is the Pierre Robin syndrome and its inclusion will therefore make a significant difference to the figures. This sub-group is also more susceptible to ascertainment bias as the prevalence of sub-mucous clefting within the general population is thought to be as common as overt isolated CP (Christensen and Fogh-Andersen, 1994). In a detailed study of isolated cleft palate in Denmark, these authors noted that there is a marked difference in sex ratios for non-syndromic overt CP including the hard palate, and non-syndromic overt CP of the soft palate only. This, combined with the tendency for hard palate and soft palate clefts not to occur within the same families, indicates that they may be two etiologically distinct sub-groups of cleft palate. Christensen and Fogh-Andersen (1994) therefore recommended that future studies on isolated cleft palate distinguish between hard palate, soft palate and sub-mucous hard palate in an attempt to disclose etiological heterogeneity within secondary palatal clefting.

The inclusion of the Pierre-Robin anomaly is also complicated by the fact that the diagnosis of Pierre-Robin is inconsistent; e.g. some clinicians insist that respiratory distress is an essential part of the anomaly while others make a diagnosis on the basis of glossoptosis and micrognathia with the cleft, whether or not there is respiratory distress.

Further complications in the consideration of isolated cleft palate are two recognized genetic phenomena:

- (a) the association of CP with 22q11.2 deletion in the velo-cardio-facial syndrome (VCF); and
- (b) X-linked clefting.

The incidence of VCF in many populations is unknown and diagnosis may be delayed, thus affecting the birth prevalence figures. X-linked clefting has been reported in some populations, such as the Icelandic population (Moore et al., 1987), but has not been investigated in many others. Also a study by Lowry and Rennick (1969), X-linked sub-mucous cleft palate that is part of an X-linked recessive trait; this might complicate the picture regarding cleft palate birth prevalence and sex ratio figures.

2.2 Recommendations for producing better descriptive statistics in OFC

2.2.1 *Population-based versus hospital-based registries*

In much of the older literature and in current work in less-developed countries, data are often available only on births delivered in hospital. Unless almost all births occur in hospital, such data may be biased. However, if hospital confinement is more available to women from the upper socioeconomic groups, hospital-derived rates may underestimate those for the community as a whole. Interpretation of hospital series, therefore, is not straightforward unless the proportion of births in the community delivered in hospital approaches 100%. Even so, when hospital records alone are searched, the number of cases expressed as a percentage of all known cases (found by using multiple sources of ascertainment) may be low, as indicated by the Hungarian figure of 52.5% based on hospital records only (Czeizel and Revesz, 1970).

While complete ascertainment is almost impossible to achieve, we can come close to it by pooling data from several overlapping sources. The quality of a population-based perinatal register will depend on how many sources are used and how thorough the ascertainment process is; also, cleft registers or hospital-based registers tend to be a subset, excluding stillbirths, early deaths, minor anomalies not requiring surgery, patients who move away, miscoding, etc. As well as being less complete, a hospital-based registry will tend to have fewer cases with associated abnormalities because of stillbirths and perinatal deaths (not requiring admission) and because another feature may be more important than the cleft.

2.2.2 *Multiple sources of ascertainment*

Multiple sources of ascertainment from population-based samples should be used for incidence statistics, and complete censuses or representative samples should be employed for prevalence statistics. These constitute the best approaches available for preparing accurate estimates of rates, because no single data source has sufficient reliability (Czeizel and Tusnadi, 1971).

In preparing incidence data to support genetic and other etiological studies, all aborted fetuses and stillbirths should either be included or appropriate adjustments made. Whether terminations and fetal deaths are included, the inclusion criteria, and the methods used should be clarified. Similarly, the effects of differential prenatal and postnatal death rates on the apparent sex ratios for clefts should be documented. All degrees of cleft expression should be diagnosed to prevent under-ascertainment.

Grouping by ethnic origin is not entirely objective

2.2.3 *Cleft-type and associated malformations*

All epidemiological and genetic data should be presented by specific cleft type whenever possible (Fogh-Andersen, 1942; Fraser, 1970). Each cleft type should be subdivided by the presence or absence of associated congenital malformations (Emanuel et al., 1973). Where possible, syndromic cleft cases should be separated from nonsyndromic ones; and the classification used and how this was done should be explained, for example, by a dysmorphologist. Birth prevalence statistics for clefts will further benefit risk-factor studies if they are tallied separately for familial and sporadic cases (Melnick et al., 1980; Bixler, 1981) in which the genetic and environmental risk factors may differ, and then for syndromic versus nonsyndromic status within these categories. Since the major cleft phenotypes are actually heterogeneous entities, disaggregating them for statistical purposes may aid the investigation of unitary disease categories.

2.2.4 *Ethnic grouping*

Where possible, data within countries should be presented by ethnic group, although it must be recognized that grouping by ethnic origin is not entirely objective. Also, in light of some emerging evidence, it may be useful to have a record of socioeconomic status. Ideally, datasets containing core information agreed by consensus should be collected while, for studies in suspected high-risk population subgroups, additional information should be collected, such as specific parental genotypes or phenotypes, older parents, medicated mothers, mothers with certain chronic diseases, and parents with unique dietary or other environmental exposures.

BOX A

Recommendations for producing better descriptive statistics in OFC and epidemiology

Orofacial clefting (OFC) is a heterogeneous group of defects with a considerable range of severity; therefore, there will inevitably be variability in ascertainment rates, and multiple sources of ascertainment should be used where possible. Studies also vary in the criteria used for differentiating syndromic from non-syndromic clefts. Many of the earlier publications were less discriminating on the differences in frequency between CP and CL/P, often quoting a combined figure. Many more recent papers do differentiate and some even subdivide CL and CLP. The validity of inter-centre comparisons is dependent on the comparison of similar groups of patients, and standardized classifications are necessary. Molecular diagnoses will increasingly assist with the differentiation and classification (*see Section 5.2*)

2.3 Conclusions

The overall conclusions to be drawn from the data presented in this chapter are as follows:

- There is ample evidence of the distinctly different nature of CL/P and CP, and emerging evidence of distinct differences in subgroups within these overall conditions.
- There is a great deal of geographical variation, more apparent for CL/P than CP.
- There is apparent variation in the proportion of OFC cases with additional congenital anomalies and syndromes.
- There is no consistent evidence of time trends, nor is there consistent variation by socioeconomic status or seasonality, but these aspects have not been adequately studied. There is a need to investigate such parameters within, as well as between, different populations.
- There is considerable international variation in the frequency of OFCs, but validity and comparability of data are adversely affected by numerous factors, among which are: source population of births considered (hospital versus population), time period, method of ascertainment, inclusion/exclusion criteria and sampling fluctuation.
- There are many parts of the world for which we have little or no information on the frequency of OFCs, in particular parts of Africa, Central Asia, Eastern Europe, India and the Middle East. This needs to be addressed urgently.

3

Possibilities for improving the treatment of CFA

Three interrelated clinical management issues were identified by participants as being priorities for international collaborative research:

- the identification and dissemination of optimal clinical interventions for the management of craniofacial anomalies (evidence-based care);
- the identification and dissemination of strategies to optimize the quality of services that deliver care (quality improvement); and
- the identification and dissemination of strategies to increase the availability of care to all affected citizens of the world (access and availability).

3.1 Evidence-based care

Evidence-based care is considered to be “the integration of best research evidence with clinical expertise and patient values”. In respect of therapeutic interventions, the most powerful evidence is derived from systematic reviews that provide a synthesis of relevant randomized controlled trials (Sackett et al., 2000).

However, for CFA care providers there are some challenges ahead. Even for the longest established CFA intervention – the management of cleft lip and palate – the scientific basis of the discipline is weak. Virtually no elements of treatment have been subjected to the rigours of contemporary clinical trial design (Roberts et al., 1991) and there is a bewildering diversity in practices. A recent survey of European cleft services revealed that, in 201 teams, 194 different surgical protocols were followed for unilateral clefts alone (Shaw et al., 2001). Table 4 shows the variation in sequence and number of operations in current use to repair a unilateral cleft in Europe.

Table 4: Sequence of operations for the repair of unilateral complete cleft lip and palate

First operation	Second operation	Third operation	Fourth operation	%
Lip closure	Hard and soft palate closure			42.8
Lip closure	Soft palate closure	Hard palate closure		15.3
Lip and hard palate closure	Soft palate closure			10.4
Lip and soft palate closure	Hard palate closure			10.0
Lip, hard and soft palate closure				5.0
Lip closure	Soft palate closure	Hard palate closure and alveolar bone grafting		3.5
Lip and soft palate closure	Hard palate closure and gingivo-alveoplasty			2.5
Lip and alveolar closure	Hard and soft palate closure			2.0
Soft palate closure	Lip and hard palate			2.0
Lip adhesion	Lip closure	Soft palate closure	Hard palate closure	1.5
Lip and alveolar closure	Soft palate closure	Hard palate closure		1.0
Lip adhesion	Lip, hard and soft palate closure			1.0
Lip adhesion	Lip and hard palate closure	Soft palate closure		1.0
Hard and soft palate closure and alveoplasty	Lip closure			0.5
Lip and soft palate closure	Hard palate closure and alveolar bone grafting			0.5
Lip adhesion	Lip closure	Hard and soft palate closure		0.5
Lip closure	Soft palate closure	Gingivo-alveoplasty	Hard palate closure	0.5
Total				100.00

Source: Shaw et al., 2001

Generally speaking, choices in surgical technique, timing and sequencing, and choices in ancillary procedures such as orthopaedics, orthodontics, and speech therapy are arrived at after disappointment in the results of former practices, rather than on the basis of firm evidence that the new protocol has succeeded elsewhere. As a consequence, the unsubstantiated testimony of enthusiasts for a particular treatment has done much to shape current practices. Typically, enthusiastic claims are made for a new type of therapy; the procedure is widely adopted; a flow of favourable anecdotal reports ensues; little or no positive evidence develops to support the desirability of the procedure; there is a sharp drop in the number of clinical reports, again without evidence to support the change (Spriestersbach et al., 1973).

3.1.1 Sources of bias in CFA research

See Box B, facing page.

Not surprisingly then, empirical research frequently demonstrates that in studies of health care interventions without randomization, an inflated view of effectiveness results (Kunz and Oxman, 1998). Thus controlled trials of a series of psychiatric medications found them effective only 25% of the time but, in uncontrolled studies of the same medications, 75% were positive. Even more dramatically, none of a series of randomized trials of portacaval shunt surgery found clear evidence of benefit but 75% of uncontrolled studies did.

3.1.2 The hierarchy of evidence for CFA research

As non-randomized studies make up the great majority of the current literature in CFA treatment they must be appraised with great caution, being appreciated for the contributions to knowledge they can make and also recognized for their inherent limitations. They conform to the following broad hierarchy (Roberts et al., 1991):

- **Anecdotal case reports:** Case reports may signal important new developments in clinical practice, but the evidence they contain for a widespread change in practice remains generally unconvincing in the absence of subsequent rigorous confirmation.
- **Case series:** Reports of a series of cases treated by the same method provide more substantial evidence of the merits of a particular technique or programme of treatment, and provide the professional

Sources of bias in CFA research

The general rules of “health technology assessment” are well established and the quality of treatment comparisons conforms to a widely accepted hierarchy, from anecdotal reports to randomized trials and systematic reviews. This hierarchy relates to the degree of effort made to minimize ever-present sources of research bias that readily lead to false conclusions. The following certainly apply to the literature concerning CFA, and make comparisons between reports unreliable:

Susceptibility bias (lack of equivalence between groups of cases): Some patients will inevitably be more susceptible to the treatment applied, because their condition is less severe or because they inherently possess a better prognosis. Thus the apparent effectiveness of any technique, applied to a group of cases that are inherently more amenable to correction, will be inflated if compared to another technique applied to a more challenging group of cases. For example, comparisons of facial growth data may be dubious where there are inherent differences in facial form between communities. Similarly, speech development may be less good in circumstances where the socioeconomic profile of the population served by a particular centre is less favourable, or where the local spoken language calls for different oro-pharyngeal skills.

Proficiency bias: In a similar manner, a more skilled surgeon or clinical team can also inflate the apparent effectiveness of a technique. If operator A is 10% better than operator B, and technique X is 5% better than technique Y, a false conclusion will be reached in a comparison of technique Y performed by A, versus technique X performed by B.

Follow-up bias: The consumer of journal or conference reports needs some reassurance that the “whole story” has been given and that follow-up has been as rigorous for the cases that went badly as for those that went well. Without knowing about all the cases on whom a particular technique was tried, reliable conclusions cannot be drawn.

Exclusion bias: In reporting the effectiveness of an intervention it is often tempting to exclude cases retrospectively, where the expected progress was not achieved. Typical grounds for retrospective exclusion might be lack of compliance on the part of the patient or suspicion that an underlying condition (e.g. an ill-defined “syndrome”) has prevented the intervention from working. Irregular application of the rules of retrospective exclusion clearly can remove any equivalence that comparison groups may have had.

Analysis bias: Given the virtual absence of agreed rating schemes for outcome evaluation, reporting in the CFA literature is inevitably inconsistent. And without objectivity in appraisal – as achieved with blinded, independent panels – comparisons must be unsure.

Reporting bias: It would appear that clinical researchers, like pharmaceutical companies, are more likely to report positive findings than negative ones. But not only are findings more likely to be reported if they are positive, but they are also more readily accepted for publication by journals, more readily accepted for conferences, more often published in English, and more often cited in later publications (Easterbrook et al., 1991; Dickersin et al., 1992; Dickersin and Min 1993; Egger et al., 1997; Stern and Simes, 1997).

Differences
arising from biases
... may exceed
actual differences
attributable to
procedures

community with a general impression of relative efficacy. Rather commonly, however, outcome is measured in the short term and the enthusiasm of the reporters may impair true objectivity. Thus primary bone grafting, first heralded as an important breakthrough in case-series reports, was later shown by randomized controlled trials to be harmful to facial growth (Rehrmann et al., 1970; Jolleys and Robertson, 1972). On the other hand, case series of secondary bone grafting using cancellous iliac crest grafts revealed persuasive evidence that one aspect of outcome, the patient's dentition, could be reliably restored beyond levels previously attainable (Boyne and Sands, 1972, 1976; Bergland et al., 1986). The immediacy of these benefits ruled against the need for a randomized trial though potential growth disturbances still deserved consideration (Semb, 1988). Future trials of bone grafting may, however, still be necessary to examine individual aspects of surgical technique or timing, or to test the suitability of alternative graft materials.

Case series rarely provide evidence of the superiority of one technique over others where a choice of broadly similar methods exists and in which any improvement may be modest rather than dramatic. This is a major problem in the evaluation of the primary surgical repair of clefts, since this may be achieved with apparently similar success by methods that differ in technique, timing and sequence. Differences arising from the biases listed above are likely to exceed actual differences attributable to the procedures.

- **Non-randomized comparison studies:** Opportunities for non-experimental comparisons of therapies or programmes of care can arise in several ways: by the coexistence of different therapies at the same centre, by the replacement of one therapy with another, or by collaboration of two or more centres. In such comparisons attempts may be made to reduce bias.

✧ **Comparison of co-existing therapies:** In using retrospective material, such as case notes or clinical databases, checks can be made on the equivalence of the groups, commonly in terms of gender, age or diagnostic subtype. Preferably, cases can be matched pair-wise on these characteristics, or adjustments can be made in the analysis by stratification or the use of multivariate statistical methods. In either case, however, doubt will remain that important prognostic factors have been masked for, if two or more therapies were being used concurrently within a single centre, selective allocation to treatment may have occurred. For example, decisions as to when (at what age) to perform surgery may be influenced by unrecorded aspects of the condition, the availability of personnel, the health of the child or

Changes in technique are often a result of changes in personnel with differences in skill

parental attitudes and characteristics. Should these factors influence outcome, confounding would occur in any study of the effect of age on surgical outcome.

Even if it is possible to match or adjust data to remove bias due to gender, age or severity, this gives no guarantee that some other prognostic factor that may affect outcome is not associated with choice of treatment. And of course, a critical factor in surgical outcome is the differing proficiency of different surgeons.

❖ **Comparison with historical controls:** These studies may arise as natural experiments by changes in therapy within a treatment centre. Such research is feasible when durable records (radiographs, study casts, speech recordings, photographs, etc.) are obtained in a standardized way for both those subjects treated by an earlier method (the historical controls) and those subjects treated by a subsequent one, allowing simultaneous evaluation. An alternative circumstance in which such studies arise is where data for a group of patients receiving a standard treatment already exists and can be gathered in a similar way when a new treatment is introduced. This design requires only half the number of patients to be gathered prospectively as a randomized clinical trial and is clearly attractive where recruitment of cases is slow. Furthermore, it has been argued that, in circumstances of poor outcome, it may be unethical to withhold new treatment in order to create a control group (Gehan, 1984).

There are nevertheless several biases and possibilities for confounding that generally tend to favour the newly-introduced procedure. In practice, changes in technique at a treatment centre often come about as a result of changes in personnel who may have performed differently in respect of the previous method. This leads to bias due to differences in skill of personnel associated with either treatment method. For example, a new method of treatment is often tested by an experienced and innovative surgeon who may be expected to achieve better results than the average surgeon. This clearly introduces the confounding effect of operator proficiency with treatment. Even where there is stability of staff, bias reflecting gradual changes of ability and technique are highly likely and definition or ascertainment of prognosis may change. New methods may also be initially applied with some selectivity to “suitable” cases as experience is gained. Other aspects of clinical management may have been altered with the intention of improving outcome, creating additional possibilities for bias in favour of the innovative procedure. Multivariate methods have been suggested as a way to adjust for these biases, but serial changes in treatment are likely to take place in parallel, resulting in a strong

Bias favouring
innovative
procedures
is a major cause
for concern
with historical
control studies

association between treatment variables (Semb et al., 1991). This is one reason why historical control design is generally unsuited to evaluating primary cleft surgery since other changes in the total programme of care are likely to have occurred during the extensive recruitment period.

The bias favouring the innovative procedure is a major cause for concern with historical control studies as they may either fail to resolve a controversy or alternatively create ethical concerns that preclude further, more rigorous, comparisons. Favourable outcomes suggested for a new procedure by historical control studies have been disputed by subsequent randomized controlled trials (Pinsky, 1984; Pollock, 1986). Thus, the danger exists that historical control studies could set in motion an unwarranted cycle of change with no benefit to the patient and consequently delay the process of development.

The reduction in recruitment time for a historical control study in which data are gathered prospectively on a new method is also less important when extended follow-up is required of each case. If, for example, the proposed follow-up of a trial of 2 methods of primary surgery is 10 years and the recruitment time of patients sufficient for a randomized trial is 4 years, the total duration would be 14 years. The potential saving of time in a partially prospective, historical control study would only be 2 years (14%).

✧ **Inter-centre comparison:** The multi-centred approach offers distinct advantages for cleft or CFA treatment centres, as the generation of adequate samples within specific subtypes treated by contrasting treatment modalities is extremely difficult. Prospectively planned recall of cases at participating centres allows data on outcome to be collected in a standardized way, and rigorous planning and execution across the centres can ensure consecutive case recruitment and consistent evaluation (Shaw et al., 1992a,b).

Provided procedures for entry into the study are equivalent in all participating centres, this strategy is extremely valuable in assessing the outcome of surgery, together with other major components of the treatment programme at respective centres. However, for primary cleft surgery it is difficult, if not impossible, to establish the key beneficial or harmful features of a specific treatment due to the invariably complex and arbitrary mix of surgical technique, timing and sequence, ancillary procedures, and surgical personnel (Shaw et al., 1992b). For example, if two centres differ in the use of presurgical orthopaedics and types of primary lip and palate surgery, there is no way to determine which of these procedures might be responsible for any difference in outcome between centres, nor would a null result

Systematic
review of
randomized trials
is the optimal
method of choice

allow the conclusion that individual aspects of the treatment programme are equivalent. The method is therefore better suited to comparative clinical audit and quality assurance than definitive clinical research. The existence of significant disparities in outcome of the overall treatment process provides a basis for speculating as to the possible cause, and inter-centre studies should, therefore, be highly motivating towards the generation of specific hypotheses for subsequent trials.

- **Randomized controlled trials:** For the comparison of therapies there is little doubt that the randomized controlled trial is generally the method of choice, scientifically and ethically. Prognostic factors, including clinical proficiency, whether known or unknown to the investigator, tend to be balanced between treatment groups. Since patients are registered prior to treatment and followed up prospectively according to a clearly defined protocol, missing data are less likely as the potential loss to follow-up and late exclusion is reduced. Formalizing the protocol at the outset, as required by an ethical review board or funding agency, increases the likelihood of impartial analysis. The likelihood of reporting the results is also increased but by no means guaranteed.

Randomized controlled trials can, of course, also be performed badly. Notably, if the randomization procedure is not strictly applied (i.e. if allocation is not fully concealed from the investigators), bias can enter. Inadequate concealment in clinical trials is associated with higher odds ratios, i.e. an inflated view of effectiveness emerges (Moher et al., 1998), as in the case of non-randomized studies. Trials with insufficient cases may also give misleading results.

- **Systematic review of randomized trials:** Systematic review of all relevant randomized trials is the optimal method for establishing whether scientific findings are consistent and can be generalized across populations, settings and treatment variations, or whether findings vary significantly by particular subsets. Explicit methods used in systematic reviews limit bias and improve reliability and accuracy of conclusions (Chalmers and Altman, 1995). Meta-analysis – the use of statistical methods to summarize the results of independent trials – can provide more precise estimates of the effects of health care than those derived from individual studies. The Cochrane Collaboration is an international organization established to prepare, maintain and promote the accessibility of systematic reviews of the effects of health-care interventions and, as randomized trials in CFA are completed and reported, it will become a primary source of reviews and dissemination (www.cochrane.org).

3.1.3 Improving the evidence base for CFA

Given the relative scarcity of CFA, the dispersion of clinical services and the diversity of therapies, the establishment of a sound evidence base seems unlikely, without the development of a strategic international framework.

Early experience with randomized trials in cleft management

Almost thirty years ago, Spriestersbach et al., (1973) identified the need for prospective research to resolve central problems of cleft management, but remarkably few randomized trials have been performed in cleft lip and palate surgery despite being the surest means of advancing the discipline in the face of overwhelming uncertainty about the relative efficacy of countless different programmes of care around the world. In a review of 25 years of the *Cleft Palate Journal*, only 5 controlled clinical trials were identified, with only 1 involving a follow-up of surgery for more than 4 years (Roberts et al., 1991).

Robertson and Jolleys conducted two small randomized controlled trials of primary surgery in the 1960s. In the first study a sample was randomized in respect of alveolar bone grafting at the time of primary surgery in infancy (Robertson and Jolleys, 1968). Follow-up revealed a detrimental effect on facial growth in the grafted group (Robertson and Jolleys, 1983). The second study involved 2 groups of 20 cases where 1 group's anterior palate closure was delayed until 5-years of age. No benefit for dentofacial growth was found in delaying hard palate closure (Robertson and Jolleys, 1974). A follow-up study when the children were 11 years of age reached the same conclusion (Robertson and Jolleys, 1990). In a quasi-randomized trial (patients entered on basis of birthdates), Wary et al. (1979) found a difference in perioperative morbidity following 3 types of palate repair in 47 patients with a variety of cleft types: V-Y pushback, Langenbeck, Langenbeck with superiorly based pharyngeal flap. Speech outcomes were subsequently reported for 52 patients (Holtman et al., 1984). Morbidity was least with the Langenbeck and speech outcomes were the same in all three. Chowdri et al. (1990) compared rotation-advancement and triangular flaps in unilateral cleft lip repair in 108 cases and found no differences in lip and nose appearance.

In another quasi-randomized controlled trial (patients alternated rather than randomized) on speech outcome, Marsh et al. (1989) compared palate repair with or without intravelar veloplasty in 51 subjects with a broad range of palatal cleft types. Speech evaluations were made at a two-year follow-up. No difference in outcome was detected but the procedure, including intravelar veloplasty, required a significantly longer operating time.

Remarkably few randomized trials have been performed in cleft lip and palate surgery

Most trials
have involved
relatively small
samples

Another randomized controlled trial on speech outcome and maxillary growth in patients with unilateral complete cleft lip and palate operated on at 6 versus 12 months of age was undertaken in Mexico (Ysunza et al., 1998). The study groups consisted of 41 subjects operated on at 12 months of age, and 35 subjects operated on at 6 months. There was no statistically significant difference in velopharyngeal insufficiency, maxillary arch development or soft tissue profile as measured on cephalometric radiographs. However, phonologic development was significantly better in patients operated at six months and none of the patients in this group developed compensatory articulation. The authors concluded that cleft palate repair performed at six months significantly enhances speech outcome and prevents compensatory articulation disorder. The same group compared minimal incision palatopharyngoplasty with and without individualized velopharyngeal surgery for velopharyngeal insufficiency in 72 patients with submucous cleft palate, and found no benefit for the more complex procedures (Ysunza et al., 2001).

For patients with velopharyngeal insufficiency (VPI), secondary surgery to the pharynx is often recommended. Whitaker et al. (1972) found no difference in outcome in a randomized trial of 35 patients, comparing superiorly- versus inferiorly-based flaps. More recently, pharyngeal flap or sphincter pharyngoplasty were compared in a multi-site randomized controlled trial of 97 patients. Patients were evaluated before surgery, then 3 and 12 months following surgery, by perceptual speech evaluation, video nasopharyngoscopy, nasometry, polysomnographic sleep study, lateral cephalometric radiographs, audiometry and tympanometry. Preliminary analysis has shown both techniques to be equally effective and equally safe (VPI Surgical Trial Group, 2001). A larger replication of this trial is currently under way at the Hospital for Research and Rehabilitation of Craniofacial Anomalies, University of São Paulo, Brazil.

Most of the above trials have involved relatively small samples, but two current surgical trials are taking place on a more ambitious scale. A randomized controlled trial to compare velopharyngeal function for speech outcomes in two groups of patients with complete unilateral cleft lip and palate is also being undertaken at the Hospital for Research and Rehabilitation of Craniofacial Anomalies in Brazil (Williams et al., 1998). The two palatoplasty techniques tested are von Langenbeck with intravelar veloplasty and the Furlow procedure. A total of 608 patients are being entered into 1 of 2 age categories; patients having surgery before 1 year of age and patients undergoing surgery at approximately 1½ years of age. This study is designed to determine which of the two surgical procedures is superior in constructing a velum capable of affecting velopharyngeal competency for the development of normal speech.

The feasibility
of randomized
trials has been
demonstrated

Since 1986, North European teams have been developing a concerted programme of multidisciplinary inter-centre research in cleft lip and palate. This includes a comparison of surgical outcome in four Scandinavian centres (Friede et al., 1991; Enemark et al., 1993) and six European centres (Shaw et al., 1992a,b; Mars et al., 1992; Asher-McDade et al., 1992; Mølsted et al., 1992, 1993a,b; Marrant and Shaw, 1996; Grunwell et al., 2000). Following these collaborations, the limitations of inter-centre studies became increasingly obvious to these teams, as it became clear that it would be impossible to separate and compare the single elements of the package of care provided in the different centres. This experience provided a compelling stimulus for starting randomized controlled trials in primary surgery of clefts and 10 centres are currently participating in a set of 3 parallel trials where groups of teams are testing their traditional local protocols against a common protocol. At the time of writing, more than half of the proposed sample of 450 infants with unilateral cleft lip and palate has been entered into this “Scandcleft” trial (Semb, 2001).

Randomized trials of other interventions have also been completed. These include a trial of artificial bone (Ping et al., 2001), a trial of nasal floor augmentation (Chen. et al., 1999), trials of anaesthesia or analgesia (Bremerich et al., 2001; Prabhu et al., 1999; Ahuja et al., 1994; Nicodemus et al., 1991), a trial of perioperative steroid therapy (Senders et al., 1999), a trial of perioperative antibiotics (Anland et al., 1995), speech therapy following velopharyngeal surgery (Pamplona et al., 1999), inclusion of mother in speech therapy (Pamplona et al., 2001), phonologic versus articulatory speech intervention (Pamplona et al., 1999), the use or non-use of presurgical orthopaedics (Kuijpers-Jagtman and Prahl, 1996; Kuijpers-Jagtman and Prahl-Andersen, 1997; Konst et al., 2000; Prahl et al., 2001), the use or non-use of arm splints following surgery (Jigjinni et al., 1993), feeding after surgery (Darzi et al., 1996; Lee et al., 1999), feeding methods in infancy (Brine et al., 1994; Shaw et al., 1999), and the use of continuous airway pressure (CPAP) in the treatment of hypernasality (Kuehn et al., in press), and fluoride supplements for dental caries (Lin and Tsai, 2000).

Such efforts demonstrate the feasibility of randomized controlled trials in the CFA field and indicate the probable shape of future progress. Thus trials of sufficient power are likely to be mounted either through collaboration between funding agencies, clinical scientists, and large, high volume centres (possibly in the developing world, as in the Brazilian trials above). Alternatively, they may be mounted as multi-centre investigations within collaborative groups with strong geographic or cultural links, as in the Scandcleft trial. Each will have a place.

History indicates that not all surgical innovations are an enduring success

Challenges in mounting clinical trials

Among the challenges in mounting clinical trials concerned with CFA are, firstly, adequate length of follow-up since interventions are often applied at an early stage of life and their full consequences only revealed some years later; secondly, the location of CFA may impair many structures and functions calling for the quantification and weighting of diverse outcomes.

Above all, however, is the challenge of sample size since the various subgroups of CFA occur infrequently. Current estimates suggest that 2 groups of around 75 cases of the same diagnostic subtype are required in trials of cleft surgery. For example, more than 1 million births would have to occur for a trial including 150 infants with complete, non-syndromic, unilateral complete cleft lip and palate (assuming a rate of 1 per 7 of all cleft types, 1 cleft per 700 births, 75% compliance with all inclusion/exclusion criteria, and consent obtained in 90% of cases). On the basis of the actual rate of entry to the Scandcleft trial mentioned above, smaller countries, such as Denmark (population 5.3 million) and Norway (population 4.4 million) would take 8 and 11 years respectively to recruit 150 cases in a single-nation trial, despite a rate of 1 cleft per 500 births.

Ethical issues in randomized trials

The ethical issues raised in randomized trials in CFA care are interesting (Berkowitz, 1995; Shaw, 1995), in particular the double standards that are applied in clinical experimentation. History indicates that not all surgical innovations are an enduring success. Discredited, though once fashionable techniques, include gastric freezing for bleeding peptic ulcer, carotid body denervation for bronchial asthma, portacaval shunt to prevent oesophageal variceal bleeding, nephropexy for viceroptosis, removal of chronically inflamed appendix and periarterial sympathectomy (Baum, 1981; Salzman, 1985). Indeed, numerous reports show that new treatments are as likely to be worse, as they are to be better, than existing alternatives (Chalmers, 1997).

Where the doctor leads, however, most patients and parents will follow, raising an important ethical dilemma. If a surgical team wishes to test an innovative procedure in a randomized trial it must obtain ethical approval from an appropriate authority and fully inform each new patient of any uncertainty and/or risk prior to obtaining his/her signed consent. Ironically, if the team wishes to try out the same innovation on all its patients, no such rules currently apply (Chalmers and Lindley, 2000). “Ethical codes that seek to protect patients ... regulate the responsible investigator but not the irresponsible adventurer” (Lantos, 1994). In the United States the National Commission for the Protection of Human

There exists a strong imperative to mount clinical trials across a range of CFA

Subjects recommended that “medical committees should be responsible for ensuring that major innovations undergo proper scientific evaluation” and be charged with “determining which new treatments need to be evaluated, the proper method of evaluation and how to limit the use ... prior to the completion of that evaluation” (Tonelli et al., 1996). As yet no such body exists, neither in the United States nor elsewhere.

In the light of the above, there exists a strong imperative to mount clinical trials across a range of CFA where true uncertainty of effectiveness (equipose) exists, and to apply the customary rules for informed consent and ethical approval from appropriate authorities. When trials in a developing country are planned and funded by a developed country, it would offer reassurance if a cooperative or parallel trial were also to be undertaken in the developed country unless, of course, the trial has relevance only for developing countries.

Planning for surgical trials

See Box C, facing page.

Measuring outcome

The ultimate goal of CFA care is restoration of the patient, as far as possible, to a “normal” life, unhindered by handicap or disability. However, the measurement of normalcy is a highly complex proposition and there is certainly no index at present that would allow sufficiently sensitive comparison between alternative treatment protocols. Clinical trials will focus more on “proximate” outcomes. These will mainly represent different aspects of anatomical form and function in the parts affected by the CFA, often reflecting the particular interests of individual provider groups. In essence, most measures will be an indication of the deficits that persist despite (or as a result of) treatment, such as shortcomings in appearance, speech, sight, hearing and dentofacial development. The general rules of reproducibility and validity apply, the latter being especially important when outcome is assessed before maturity. Longitudinal archives may be useful to determine the reliability of prediction for outcomes that are to be measured in the young (Shaw and Semb, 1996; Atack et al., 1997).

Meaningful ways to document the satisfaction of patients and their families are essential, but present scales are rudimentary and may possess little validity. The development of techniques that have cross-cultural international validity has not begun and will be a significant challenge.

In relation to cleft surgery, experience with a number of outcome measures and scales have been obtained regarding speech, dentofacial outcomes and patient satisfaction (e.g. Kuehn and Moller, 2000; Sell et al., 2001; Williams et al., 2001). Further work is certainly needed to refine these and build

Systematic planning for surgical trials

Whereas hypotheses for clinical trials in many disciplines will frequently be generated by laboratory-based studies or a consideration of previously reported cohort studies and clinical trials, this is unlikely to be the case for surgical trials in CFA surgery, at least for some time. Animal studies can shed some light on the general consequences of scars in the palatal mucoperiosteum, for example, but inferences for human maxillary growth are questionable (Kremenak, 1984; Friede, 1998; Leenstra et al., 1999). Furthermore, speech, a key outcome for cleft surgery is a uniquely human behaviour. The opportunity for most surgeons to gain meaningful experience of different techniques is severely constrained by the relative rarity of CFA subtypes, the need for lengthy follow-up, and the lack of robust measures of outcome. Together with the probable biases that apply to the existing CFA literature, research planning may be very idiosyncratic.

In the absence of relevant animal studies and reliable clinical studies a process of informed negotiation would assist in defining promising alternatives in CFA surgery and in achieving the equipoise that must be established if clinicians are to enter ethically-grounded trials. By further negotiation, variations in current practices among potential partners could be harmonized/rationalized to create more manageable aggregations of trialists. One solution would be adoption of a focus group process supported by literature review specialists. Members of the focus groups would be selected on the basis of their knowledge and experience in the field, and their standing; the latter to encourage maximum credibility of the process and foster wide implementation of eventual trial findings. They would also be selected on their likely willingness and ability to enter and/or recruit surgical centres for the eventual trial. Collectively the focus groups should represent a good geographic and multidisciplinary spread.

For different clinical topics such a process would define promising therapies, appropriate outcome measures, randomization schemes, and potential partners to develop cooperatives and funding applications.

consensus upon international standards. Reliable rating of appearance is still problematical and, for speech, linguistic differences represent a significant international challenge. Outcomes should be patient-centred, i.e. measuring things that matter to ordinary people, rather than sophisticated surrogate measurements that may have little relevance to everyday life.

Indeed, measurements of aesthetic and functional outcomes in isolation are not good predictors of emotional (psychological) adjustment and well-being (Robinson, 1997). There is a pressing need to identify the variables

that contribute to the quality of life of affected individuals. Once identified, this knowledge should then be used to develop and refine methods of support and intervention, designed to optimize psychosocial as well as aesthetic and functional outcomes in CFA.

BOX D**Measuring treatment burden**

Since the consequences of CFA may be apparent through every phase of childhood and adolescence, there is seldom a time when the disciplines involved in care cannot recommend one or another intervention. The powerful desire of patients and parents to reach the point where the stigma of CFA will be completely eradicated makes it likely that they will accept most proposals and willingly comply with protocols of care recommended by all members of the team, no matter how demanding they may be. They have little choice.

So far, "burden of care" has received little attention in CFA studies, yet the combined total of operations – other treatment episodes, and review appointments for the first 20 years of life, including all the disciplines that may be involved – can be enormous. Apart from pain and suffering and the disruption to family life, employment and school attendance, the dependent role in which this places the patient may have an adverse effect on the patient's sense of self-determination or locus of control.

A particular problem has arisen over the years with supplementary orthodontic interventions such as presurgical orthopaedics, primary dentition orthodontics and maxillary protraction. There is little evidence to suggest that the extra burden imposed on patients and the financial cost of these interventions is justified by any significant benefit (Severens et al., 1998; Long et al., 2001). Thus it is important in clinical trials to accurately record the total number of ancillary interventions and clinical visits in addition to surgical episodes.

Measuring cost-benefit

Economic pressures around the world have forced close examination of the true financial costs of treatment and, with reducing budgets, clinicians must either be involved in cost controls or have arbitrary choices imposed upon them. Surgical operations are invariably expensive treatment episodes and successful initial operations that minimize the need for multiple secondary revisions are highly desirable. Furthermore, successful initial repairs are likely to reduce the duration and complexity of subsequent ancillary procedures.

Work has yet to begin in applying the techniques of health economics to the field of CFA. Health status and the utility of care and associated quality

of life may be estimated using the techniques of time trade-off and conjoint analysis (Torrance, 1976; Ryan et al., 1998; Ryan, 1999).

Economic prioritization models use decision analysis and simulation to assess the resource costs and patient benefits of current treatment patterns and the “cost-effectiveness gap” or potential gain from alternative surgical procedures for CFA. This would include reviews of existing literature, observational and audit databases to determine: the natural history of CFA; the incidence and prevalence of CFA; the possible indications and target populations for surgery; current treatment patterns and relevant comparators; and the costs and benefits of current treatment.

Prospective registries occupy an intermediate position between non-randomized studies and randomized controlled trials

Prospective registries – a preliminary approach for rare and/or novel interventions

During the introductory phase of a new therapy it may be impossible to mount a randomized trial if the intervention is undergoing constant modification and the population it is applied to is heterogeneous and ill-defined. Such is currently the case with many CFA interventions. A case in point in the last decade is distraction osteogenesis (gradual mechanical elongation of a bone) in its increasing application to the craniofacial skeleton.

Pending the conduct of clinical trials, the establishment of prospective registries to enable critical appraisal of different kinds of CFA interventions will maximize collective experience and minimize the biases that inevitably occur with ad hoc reporting. Such registries would therefore play a similar role to Phase I trials of pharmaceutical interventions. One such registry has been set up for distraction osteogenesis in Europe as part of the EUROCRAN programme, with centres submitting duplicate records prior to – as well as after – treatment, as a step to minimizing follow-up, analysis and reporting bias (www.eurocran.net).

As records of all cases would be filed with the registry prior to the start of treatment as well as after it, justification for non-follow-up would be required. And, as in well-conducted clinical trials, analysis bias could be overcome by employing blinded independent raters, while reporting bias could be overcome by the greater impartiality of the partnership and its predetermined conventions. Susceptibility bias and exclusion bias could not be minimized with the assurance derived from random allocation, but some checks of equivalence might be possible. Clinical proficiency, however, would inevitably remain as a major bias. Thus, prospective registries occupy an intermediate position between non-randomized studies and randomized controlled trials.

The registry approach will maximize opportunities for preparatory work on outcome methodology: for early detection of extremely promising or unpromising clinical strategies, for defining answerable questions amenable to clinical trials, and for building the interpersonal trust and institutional partnerships that will be necessary to mount such trials.

3.1.4 Tissue engineering

Tissue
engineering offers
two generic
approaches
to assist
reconstruction

Surgical advances of a more general, fundamental nature hold promise for improved CFA surgery in the foreseeable future. The discovery that, for example, wounds incurred during early gestation heal perfectly with no scars has led to intensive research of the cellular and molecular differences between scar-free healing and scar-forming healing (Whitby and Ferguson, 1991; Shah et al., 1992, 1996; Ferguson et al., 1996; Cornelissen et al., 2000a, 2000b, 1999a, 1999b). Thus the identification of high levels of TGF β 3, with low levels of TGF β 1 and 2, in scar-free wounds has led to the development of pharmaceutical interventions to reduce scarring in experimental skin wounds (e.g. www.renovo-ltd.com). Such interventions are currently undergoing trials in human volunteers and could offer considerable therapeutic benefits in surgery for cleft lip and palate and other CFA.

A major problem in the surgical treatment of CFA is the deficiency of tissue available for surgical repair – bone, muscle, mucosa or specialized dental or eyelid tissues. Tissue engineering offers two generic approaches to assist reconstruction: either to grow cells outside the body, usually harvested from biopsy specimens, or to apply some form of scaffold to orientate the repair potential of the patient's own cells *in situ*. Both approaches can be combined and it is now recognized that many of the cells participating in repair processes are stem cells, derived principally from bone marrow.

Sophisticated scaffolds can be custom-made for the individual patient by defining the anatomical defect through three-dimensional reconstruction of CAT scan and MRI images and linkage to a prototyping or milling machine to manufacture a scaffold for the precise defect. Even the most delicate microsurgery is unable to accurately restore the muscle deficiencies of clefts of the lip and palate, but there is the prospect of encouraging muscle growth along a template of the body's own proteins or a bio-degradable polymer. Signalling by growth-factor release will enhance migration.

Biomaterial science offers a potential solution for certain mechanical problems in CFA. Bone distraction techniques are effective in inducing bone formation and may be combined with osseointegration devices to allow longer-term movements of hard tissues. Detailed knowledge of

internal stress analysis can be combined with cellular reactions to force-mechanotransduction to provide information to direct growth and tissue movement.

The establishment of experienced clinical trial cooperatives will be essential to the safe, efficient and critical translation of these technologies into common practice.

3.1.5 *Research on treatment*

BOX E

Priorities for research on treatment

There is an urgent need for the creation of collaborative groups in order to assemble a critical mass of expertise and to sufficiently access large samples of patients for adequately-powered clinical trials.

Given the currently poor state of evidence for virtually all aspects of clinical management, there is an almost unlimited list of trials that could be initiated. However, the following were considered to be especially important:

- trials of surgical methods for the repair of different orofacial cleft subtypes, not just unilateral clefts;
- trials of surgical methods for the correction of velopharyngeal insufficiency;
- trials of the use of prophylactic ventilation tubes (grommets) for middle-ear disease in patients with cleft palate;
- trials of adjunctive procedures in cleft care, especially those that place an increased burden on the patient, family or medical services, such as presurgical orthopaedics, primary dentition orthodontics and maxillary protraction;
- trials of methods for management of perioperative pain, swelling and infection; and nursing;
- trials of methods to optimize feeding before and after surgery;
- trials addressing the special circumstances of care in the developing world in respect of surgical, anaesthetic and nursing care;
- trials of different modalities of speech therapy, orthodontic treatment and counselling.

Equally urgent is the need to create collaborative groups, or improve the networking of existing groups, in order to develop and standardize outcome measures; there is an especially urgent need for work on psychological and quality of life measures, and economic outcomes.

For rarer interventions, prospective registries should be established to hasten collaborative monitoring and critical appraisal, equivalent to Phase I trials. Relevant topics would be craniosynostosis surgery, ear reconstruction, distraction osteogenesis for hemifacial macrosomia and other skeletal variations, midface surgery in craniofacial dysostosis, and correction of hypertelorism.

3.2 Quality improvement

Previous research demonstrates that similar interventions achieve widely different outcomes dependent upon the manner and circumstances in which care is provided. For example, secondary complications have been found to occur up to 10 times more frequently when the care of children with unilateral cleft lip and palate is performed inexpertly or delivered in an uncoordinated manner (Bearn et al., 2001). It is evident, too, that simple care can achieve equivalent or superior outcomes to complex care at less human and economic cost (Shaw et al., 1992b; Severens et al., 1998).

The exploration of methods to define attainable standards of care for CFA and to promote quality-improvement protocols among the providers of care was considered to be an important priority.

3.2.1 Organization of services

Delegates discussed the programme of quality-improvement activity conducted under the auspices of the European Commission between 1996-2000 (Shaw et al., 2001). This activity revealed great variability between countries in the provision of medical services for individuals with cleft lip and/or palate. While long-standing high-volume centres of expertise prevailed in Scandinavia, countries such as Italy, Germany, Switzerland and (until recently) the United Kingdom, provided cleft care via large numbers of local services with small case-loads. In other countries, such as Greece, Portugal and Spain, the concept of comprehensive specialist-team care was still undeveloped.

The challenge of improving services in a pan-European manner was addressed in part by the consensual development of clinical and organizational guidelines. The difficulties observed in configuring services into specialized units with sufficient case-loads to foster proficiency of care and secure adequate resources for comprehensive care were by no means solely economic. Instead, the obstacles were frequently reported to be:

- personal egotism of individuals unwilling to discontinue the practice of treating a few children each year;
- competition between specialities for pre-eminence in the field e.g. plastic versus maxillofacial versus paediatric versus ear, nose and throat (ENT) surgery;
- local pride, with every hospital, town or region desiring its own small team;
- lack of clinical leadership; lack of responsiveness of the health authorities at local and national level.

There is
great variability
between countries
in the provision
of medical services

It was also noted that all the above problems had confronted the United Kingdom in the recent past and were not resolved until a national review was instigated by a government body (Sandy et al., 2001). The review included a national survey that revealed that Britain's fragmented, decentralized services were achieving a low standard of clinical success. As a result the government instructed regions to provide care from a single regional centre, with a fully comprehensive specialist team – typically with two to three surgeons – each responsible for not less than 40-50 new personal cases requiring primary surgery per year. In this instance, government intervention was essential to the improvement of services when voluntary methods failed (Sandy et al., 2001).

Elsewhere in Europe it was noted that the consensual guidelines on policies, practice guidelines and record-keeping had also been a powerful force in promoting reorganization of services for orofacial clefts, suggesting the influence of peer pressure at a national level. Thus within months of the publication of the European guidelines, more than half the countries in Europe had reconfigured services, formed new multidisciplinary collaborative associations, or increased funding for clinical services (Shaw et al., 2001).

3.2.2 *International recommendations*

BOX F

International recommendations on organization of cleft lip and palate services

Delegates discussed the desirability of global recommendations on the principles that should govern clinical services for clefts of the lip and/or palate, and concerning basic clinical record collection. It was concluded that such guidelines would improve clinical research capability, and also encourage improved clinical care. There was special recognition of the economic constraints that would be faced by developing countries in complying with generic guidelines, but it was felt that these were still desirable to serve as a long-term goal.

In particular, a set of guidelines recently developed through international consensus in Europe was reviewed. Delegates felt that these were appropriate as a basic requirement for wider international use and that the protocols recommended for clinical record collection were also acceptable as a minimum requirement. The recommendations of the WHO consensus conference are set out in Section 8.

The rationale for recommending case-loads of 40 or more cases per operator is largely one of statistical imperatives: comparative clinical audit and research require adequate samples of cases with a similar prognosis. Clefts of the lip and palate present with great heterogeneity, and the only

substantial category that is reasonably homogeneous is non-syndromic unilateral cleft of the lip and palate (UCLP). Even this group has considerable between-case variation, and reasonably large samples are required for statistical comparison. The Euroleft Report (Shaw et al., 1992a) provided estimates of the sample sizes required to detect differences for a variety of outcomes. The Goslon Score, a rating of dental arch relationship (Mars et al., 1987) was found to require the lowest sample size for discerning differences among groups. One half point on the Goslon scale was the extent of the differences between the top- and middle-ranked centres and between the middle- and bottom-ranked centres in the Euroleft study, equating to a 20% difference in osteotomy rate among such centres. At 5% probability and 80% power, detection of a 0.5 Goslon scale point difference in 10-year olds requires samples of the following size:

- 42 UCLP cases required in a 2-group comparison;
- 63 required in a 5-group comparison with 1 standard; and
- 77 required in a 6-group mutual comparison.

Based on an occurrence of one non-syndromic complete unilateral cleft of the lip and palate, per six clefts of all types, Table 5 (below) shows the time it would take for surgeons, with a differing annual volume of cleft work, to generate varying samples.

Table 5: Years required for the generation of samples of UCLP, related to case-load

Surgeon volume	Years to accrue sample for comparison		
	2-group comparison (n = 42)	5-group versus standard (n = 63)	6-group mutual comparison (n = 77)
6 cases per year	42	63	77
30 cases per year	8	12	15
60 cases per year	4	6	7.5

Source: Beam et al., 2001

Even if follow-up is restricted to 5 rather than 10 years or more, it is clear that only operators treating 60 new cases per year would be able to audit their outcome within a decade. In the case of the United Kingdom, the figure of 40 cases per year (requiring approximately 12 years for an audit cycle) was the compromise reached.

3.2.3 Monitoring outcomes

Participants agreed upon the desirability of establishing international standards, such as the development of rating methodology and sample-size estimates for comparison studies in the procedures of outcome evaluation, a process that also has a research dimension. Currently two general approaches were identified:

- **Inter-centre comparisons:** These might take the form of blinded comparison of records of consecutive cases from different centres, a number of which have been reported (*see Section 3.1.2*). Alternatively, one set of records may be compiled to serve as a standard reference archive against which any team could compare its outcomes. A “good practice” archive of this kind might include durable records such as study casts, radiographs, speech tapes and so forth that would be representative of the ethnic population treated by well established teams with consistent protocols. Other teams could measure their own outcome records against these. In time a series of such archives for clefts and other CFA from different regions could become a web-based resource. The development of such an archive for Europe is included in the EUROCRAN programme (*see Annex 2*).

In either case the recommended timetable for record collection would be helpful to maximize the opportunity for teams to successfully match their records to those from other centres (*see Annex 5*).

- **Registries:** Under the auspices of the American Cleft Palate-Craniofacial Association, a web-based “Craniofacial Outcomes Registry” (COR) was recently established, enabling North American teams to anonymously enter diagnostic and outcome data. Teams rate their own outcomes and can obtain an indication of their relative success compared with the Registry’s aggregated data (www.cregistry.org).

A national registry for the Cranofacial Anomalies Network in the United Kingdom has also been established and is developing protocols for standardized outcome data collection (www.perinatal.org.uk/crane).

The Swedish Cleft Palate Association also has a web-based registry (Swedish National Quality Registry for Cleft Lip and Palate Treatment, <http://natqa.uas.se/LKGreg/LKGreg.ihtml>). It is intended that teams will display the actual records of consecutive cases, allowing peer review by each other.

Participants in the meeting considered that joint, international work, in an effort to harmonize these differing approaches, was urgently required.

Only operators treating 60 new cases per year could audit their outcome within a decade

3.3 Access and availability

The meeting's attention was drawn to the fact that, by the early 1960s, most industrialized countries had gained control of diseases caused by infection and/or malnutrition, and genetic disorders and birth defects had attained public health significance (Christianson, 2001). This situation is considered to occur when the infant mortality rate (IMR) falls below 40-50/1000 live births, at which juncture countries tend to recognize the need for medical genetic services. Approximately 40 years later, a significant proportion of the world's developing nations has attained a similar situation: in 1997, 75 (53%) of the developing world's countries, in which 60% of their population resided, had an IMR of less than 50 per 1000 live births.

For the majority of CFA-affected individuals, there is a full life expectancy

Only a minority of CFA are lethal and, for the majority of affected individuals, there is a full life expectancy. Appearance, function and social integration can, in nearly all cases, be improved by surgery and related multidisciplinary specialist medical care. The cost of treatment through infancy, childhood and beyond can be considerable however and, in the developing world, often unaffordable.

For example, in 1994, the medical costs of one individual with cleft lip/palate in the United States was estimated at US\$ 101 000 (Waitzman, 1994). In the United Kingdom, the estimated cost of 1 regional multidisciplinary cleft lip and palate service, receiving 140 new cases annually, is UK£ 6.4 million per year, excluding capital costs (National Health Service, United Kingdom, 2001). The social costs of unmet or partially-met medical needs are also enormous. Affected individuals are liable to suffer stigmatization, social exclusion and barriers to employment.

When malnutrition and communicable diseases represent more pressing priorities, CFA care provided by nongovernmental organizations (NGOs), through charitable missions of medical staff or the external sponsorship of local providers, may be the only chance of treatment many individuals will have. Such efforts are known to be taking place on a remarkably large scale and in a wide variety of ways. Because of the distinctive features of these services it was considered that particular research questions need to be addressed in order to maximize the benefit of NGO endeavours in CFA. For example, in developing countries, patients often present for surgery at later ages than in developed countries, the services themselves may be of a rudimentary nature, and patients may be seen only once. Thus, a sound evidence base is needed to maximize effectiveness, safety and capacity. Again, quality-improvement strategies should be considered alongside this.

Agencies ...
should be
encouraged to
develop and
adhere to a
common
international
code of practice

3.3.1 Main approaches

Three main approaches to the provision of specialist care in the developing world were noted. The first was the establishment of efficiently run, high volume, indigenous centres of excellence, capable of serving large and widespread populations via a mixture of assisted travelling arrangements and outreach satellites. An example of such a centre that had achieved considerable success, both in providing service and conducting research, was presented (www.centrinho.usp.br).

Secondly, some NGOs assist large numbers of individuals to receive surgery by providing financial support for indigenous clinical units to undertake operations that could not otherwise be afforded. Support for training indigenous specialists may also be provided (e.g. www.smiletrain.org).

Thirdly, a large number of NGOs provide care by forming surgical missions where teams of surgeons and ancillary staff make visits to selected sites where there is a shortage of resources or experienced personnel (e.g. www.operationsmile.org; www.rotaplast.org). In several instances valuable research, especially of a genetic or epidemiological nature, has been conducted alongside these ventures (Lidral AC et al., 1997; Murray JC et al., 1997).

Ethical issues are a prominent concern in this work and some programmes have been criticized on grounds of safety, surgical competence and absence of follow-up. Though not a research issue *per se*, it was felt that the present research programme taking place under WHO auspices should attempt to encourage agencies involved in the charitable provision of treatment in the developing world to develop and adhere to a common international code of practice. Such an effort might build upon the survey undertaken by an earlier international task force on volunteer cleft missions (Yeow et al., 1997).

3.3.2 Further work

Participants identified several areas deserving further work:

- a survey of the charitable organizations involved and the scale of their work;
- an appraisal of the cost-effectiveness and clinical effectiveness of the different models of aid;
- the promotion of dialogue between different NGOs to develop commonly-agreed codes of practice and adoption of the most appropriate forms of aid for local circumstances, with an emphasis on support that favours indigenous long-term solutions;

- the initiation of clinical trials concerning the specifics of surgery in a developing country setting: one-stage operations, optimal late primary surgery, anaesthesia protocols (e.g. local anaesthetic, inhalation sedation, antisepsis);
- the development of common core protocols for genetic, epidemiological and nutritional studies alongside surgery.

3.4 Regional perspectives

The membership of the meeting was not intended to be fully representative of all nations. Several general observations, however, are possible, based upon the information presented.

A regional
“good practice”
reference archive
for sub-Saharan
Africa would be
valuable

Africa: In sub-Saharan Africa clinical resources for CFA are scarce as a consequence of prevailing economic problems and the greater challenge of communicable diseases, particularly AIDS. For example, in Namibia despite a high reported incidence, there are no cleft surgeons. As the wealthiest sub-Saharan country, South Africa has around 12 centres that undertake cleft surgery but these tend to work independently without common quality-improvement protocols. There has, as yet, been little formal study of CFA in the African population of sub-Saharan Africa and a regional “good practice” reference archive for this region would be valuable.

There are a number of centres in the cities of Northern Africa but, as elsewhere in Africa, a survey has yet to be undertaken to identify potential sites with capability for collaborative research.

Australia and New Zealand: There are well-developed services in many cities, though in some instances, the case-load is quite low, limiting the potential for collaborative research. However, the establishment of the Australian and New Zealand Craniofacial Association makes coordination possible and one centre has a programme of support and development for Indonesian and Malaysian cleft centres.

China: In China there is reportedly a high level of unmet need for cleft and other CFA treatment. There is, however, a network of several large surgical centres that could form a potential research partnership.

Treatment, however, is not free and follow-up is difficult. Speech therapists are especially scarce. Of those individuals receiving cleft surgery, only 30% are operated in the first year of life. Again this points to a need for surgical trials to define preferred operative techniques in more mature patients. A survey of clinical services and potential collaborating sites would be valuable, as would development of a quality-improvement strategy and “good practice” archive.

The Indian
subcontinent
has numerous
potential partners
for clinical trials

Europe: European clinical services have recently been surveyed (Shaw et al., 2001). In the main, Europe's problems arise from fragmentation of care over numerous small centres. The adoption of consensus recommendations, however, has begun to bring about restructuring, at least for cleft services. Several international research collaborations are under way (*see Annex 1*) and, under the EUROCRAN programme that was initiated in 2001, the European Commission is funding a series of multinational work packages that would be capable of wider networking (*see Annex 2*).

Indian subcontinent: As yet the subcontinent has not been surveyed regarding CFA or cleft services and research capability. However, an overview of India was presented and may be reasonably representative of adjoining countries. There are high levels of unmet needs and access is complicated as the majority of the population live in rural communities. There are several hundred surgeons trained in cleft surgery and several large university hospitals but, as yet, no quality-improvement protocols are in place. The subcontinent undoubtedly has numerous potential partners for clinical trials though resourcing follow-up studies will be a challenge.

Latin America and the Caribbean: As yet no survey has been done on clinical services and research capability across the continent. Mexico was represented and has at least one large centre that has successfully completed clinical trials (Ysunza et al., 1998, 2001; Pamplona et al., 2001), and is recognized as a centre of excellence in the region. Brazil was also represented by the centre of excellence at Bauru. Elsewhere in Latin America there is undoubtedly a high level of unmet need.

Sout-East Asia: Singapore has already embarked upon a surgical trial in collaboration with a large centre of excellence in Taipei (www.nncf.org; www.cgmh.org.tw) and together they have a high research capability. In Indonesia there are high levels of unmet need but around six cleft teams are established and would be potential sites for research collaboration. Already both Indonesia and Malaysia are engaged in epidemiological, nutritional and genetic research with agencies in Australia, Europe, Singapore and elsewhere. There are reportedly high local incidences of CFA, such as frontal encephalocele, that may be fruitful targets for multidisciplinary research.

Like Europe, Japan may have a fragmentation of services in small centres; however, the Japanese Cleft Palate Association has begun discussions on inter-centre studies and clinical trials. In Korea, several high-volume centres are potential sites for collaborative research and the Korean Cleft Palate Association has begun discussion on inter-centre studies.

Middle East: A high level of unmet need has been reported with few established CFA centres. A number of university hospitals in the region would be potential partners in research.

North America: North America also suffers from a fragmentation of cleft and craniofacial services, and representatives from there spoke of the difficulties of obtaining sufficient subjects for clinical trials because of the decentralized nature of services. The recent emergence of health management organizations was seen as a particular force for the fragmentation of services and dissipation of established cleft teams. None the less, the Childhood Cancer Study Group has achieved a high level of coverage in the United States, as a result of which a high proportion of affected children are enrolled in trials (Ross et al., 1996; Shocat et al., 2001).

The American Cleft Palate-Craniofacial Association has promoted adequate team care and has published several sets of guidelines, as well as initiating the Craniofacial Outcomes Registry.