EuroQol Working Paper Series

ORIGINAL RESEARCH

28 Years of the EuroQol Group: An Overview

Richard G Brooks

Abstract

This paper presents an overview of the history of the EuroQol Group from its inception in 1987. An initial outline frames the history in terms of the early formative years 1987-90, the expansion of activity in the 1990s, the Biomed EQ-net Project 1998-2001, the period 2002-2009, and post-2009 developments. After an overview of the formative years, the paper proceeds thematically to cover: the descriptive system of the instrument, termed 'EQ-5D' from 1995; translation, language and semantic issues; valuation; applications; and product development and research strategy. The material on the 3L descriptive system covers the choice and number of dimensions and the number of levels. The construction of the EQ-5D-5L version is detailed and the concept of bolt-ons is introduced. The translation section explains how the Group managed the increase in the number of language versions, and how it dealt with issues such as semantic equivalence which emerged during the translation process, with particular emphasis on the detailed work accomplished in the EQ-net project. The valuation section is lengthy, reflecting the importance to the Group of providing valuations for the health states defined by the EQ-5D instrument. The material is largely tackled chronologically, from the early 3L work through to the multi-national programme which delivered valuations for EQ-5D-5L. Topics covered include, *inter alia*, the York Measurement and Valuation of Health (MVH) programme, the EQ-net project, the Paris Protocol, the development of a crosswalk from 3L to 5L valuations, and the multi-national programme and consequent 5L protocol. The applications section details the many and varied uses of EQ-5D. The product development and research strategy section provides the opportunity to cover a number of important features of the group's work: digital strategy; the activities of members of the Youth Task Force, whose efforts delivered the 'child-friendly' EQ-5D-Y; the work of the Population Health Task Force; and the structure of Working Groups set up in 2012/13, to be re-shaped in 2015. The paper concludes with a summary of the main issues which have characterized the work of the Group.

Richard G Brooks

Content

Abbreviations

AQLQ	Asthma Quality of Life Questionnaire
BTD	Better than dead
CAPI	Computer-assisted personal interview
COPD	Chronic obstructive pulmonary disease
CREATE	Checklist for Reporting Valuation Studies of the EQ-5D
CSPTF	Condition-Specific Task Force
DAWG	Data Analysis Working Group
DC	Discrete choice
DCE	Discrete choice experiments
DEVT	Development Electronic Valuation Technology
DTF	Digital Task Force
EC	European Commission
EEP	EuroQol Electronic Programme
EQ	EuroQol
EQ-5D	EuroQol 5 dimension instrument
EQ-5D-3L	EQ-5D 3 level instrument
EQ-5D-5L	EQ-5D 5 level instrument
EQ-5D-Y	EQ-5D youth instrument
EQ-5D-5L-Y	EQ-5D-5L youth instrument
EQ-VT	EuroQol Valuation Technology
FEDEV	Further Exploration of Discrete Choice Experiments with duration for EQ-5D Valuation
HIV/AIDS	Human immunodeficiency virus infection and acquired immune deficiency syndrome
HRQoL	Health-related quality of life
HTA	Health technology assessment
HUI2, HUI3	Health Utilities Index versions 2 and 3
IVR	Interactive voice response
LHSA WG	Large Scale Health Applications Working Group
LT-TTO	Lead-time TTO
ME	Magnitude estimation
MVH	Measurement and Valuation of Health project
NHS	National Health Service
NICE	National Institute for Clinical Excellence
PDA	Personal Digital Assistant

PHTF	Population Health Task Force
PRET	Preparation for the Re-valuation of the EQ-5D Tariff
PRET-AS	PRET Additional Sample
PROMs	Patient-Reported Outcome Measures
PROMIS-43	Patient Reported Outcomes Measurement Information System 43-item short form
PTO	Person trade-off
QALY	Quality-adjusted life-year
SF-36D	Short Form 36 dimensions
SF-6D	Short-Form 6 dimensions
SG	Standard gamble
SOPs	Standard operating procedures
TTO	Time trade-off
UK	United Kingdom
USA	United States of America
VAS	Visual analogue scale
VMC	Version Management Committee
VMG	Version Management Group
VMWG	Valuation Methodology Working Group
VSWG	Value Sets Working Group
VTF	Valuation Task Force
WG	Working Group
WTD	worse than dead
YTF	Youth Task Force
3L	3 level
5L	5 level
5D	5 dimensions
15D	15 Dimension instrument

1. Introduction

This paper is a condensed and updated history of the EuroQol Group following the 25-year history presented in book form [1]. The main aim is to provide a concise history of the Group's activities whilst preserving enough detail of its deliberations to show how much work was put into developing, maintaining, and taking forward a health status instrument.

Initially an overview of the Group's activities is provided. This covers the early years 1987-90, expansion in the 1990s, the Biomed EQ-net project, the period 2002-09, and post-2009 developments including the latest phase of Working Groups (WGs).

There follows a more detailed treatment of the initial development of the EuroQol instrument, and then the construction of the descriptive system is examined. After more than a decade's experience with the instrument – termed 'EQ-5D' in the mid-1990s – the Group explored an increase in the number of levels from three (3L) to five (5L), deploying an Increased Level Task Force for this purpose. The resultant EQ-5D-5L was promulgated by the Group in late 2008. A further development with respect to the descriptive system was the concept of 'bolt-ons' in the form of either disease-specific 'tails' to EQ-5D, or additional dimensions: these are briefly treated.

Critical to the Group's operations have been translation, language, and semantic considerations. The expansion in the number of EQ-5D language versions has conformed to international translation norms, overseen by the Translations Committee, later the Version Management Group (VMG), as modes of delivering EQ-5D increased. Accompanying all this activity was feedback into the wording of EQ-5D in the English source version. The opportunity was taken in the EQ-net project (1998-2001) to investigate translation issues in detail: the paper reflects this. Finally this section points to language matters with respect to the youth instrument EQ-5D-Y, and to EQ-5D-5L, with the role of the VMG outlined.

The next section examines in some detail the valuation aspects of the Group's work. Regarded as an important objective from the first meeting, valuation of EQ-5D health states has taken up a significant proportion of the Group's efforts, with considerable resources devoted its pursuance. It is a lengthy section whose introduction outlines the topics and issues raised.

The subsequent section is devoted to applications of EQ-5D in a variety of contexts. This reflects the aim of the Group that its instrument should be of relevance in decisions made with respect to people's health-related quality of life (HRQoL).

The history of the Group is completed by a review of the product development pursued by the Group with its associated research strategy. This permits consideration of digital strategy, the work of the Youth Task Force (YTF) and Population Health Task Force (PHTF) and the implementation of a set of WGs. The structure of these WGs is being changed as a new Strategic Plan is being pursued in 2015. It is worth noting in this context that the EuroQol Group Foundation (promulgated in 1995) changed its name to the EuroQol Research Foundation in October 2014, to better reflect the nature of the Group.

This condensed history omits formal consideration of EuroQol Group organisation and administration, the context in which the activities described in this paper took place. These aspects of the Group activities are detailed in the history volume, which covers, *inter alia*: the organisation of Plenary Meetings; the composition, roles and activities of the Executive Committee and Board; and Group membership issues [1]. The substantial input from Group members in conjunction with staff of the Executive Office in Rotterdam, the Netherlands, on these matters, has ensured that the EuroQol Group has been in a position to undertake the considerable array of activities outlined in this paper.

2. Outline of Group's history

The 28-year history of the EuroQol Group can be usefully divided into a number of time periods: the early years 1987-90, 1990s expansion, the Biomed EQ-net project 1998-2001, the period 2002-2009, and the recent era. Before detailing in subsequent sections the key aspects of the Group's work over the years a brief outline based on this framework is provided.

2.1 The early years 1987-90

The first meeting of the people who developed the EuroQol instrument took place in Rotterdam in May 1987 and was designated the *Informal Workshop on Measurement and Valuation of Health-Related Quality of Life*. The 2nd (October 1987), 3rd (January 1988), and 4th (July 1988) meetings were all held at Brunel University, UK, with the group now called *The European Common Core Group*. The 5th meeting was again held at Brunel, in July 1989, and the group became *The EuroQol Group*. The 6th meeting was held in Rotterdam (January 1990), and the 7th in York (September 1990). The latter meeting was of considerable significance as the features of what was termed 'the EuroQol instrument' had finally been put in place. The instrument had been simultaneously developed in English (the 'reference' language), Dutch, Swedish, Finnish, and Norwegian.

2.2 1990s expansion

The first Plenary Meeting was held in Sweden in 1991. Papers presented were published as Plenary Meeting proceedings, with the 'business'of the Group dealt with in separate minutes. This basic format has endured with some variations to the present day. Gradually the numbers attending the meetings expanded and, in particular, members came on board from other countries with their own language versions. The first new arrivals were Spanish and Catalan, followed by French and German. By 2002 a further 63 language versions had been produced [2]. In addition the instrument was beginning to be used more widely, not least by medical personnel keen to measure HRQoL. The first paper to reflect this development concerned stroke [3] with a further paper, on cancer, published later that year [4]. After 2 further medical papers in 1994 and 1995, numbers reached double figures in 1996, and expanded considerably thereafter (see section 7).

It was becoming clear that these developments, and the copywriting of the instrument, meant that the relatively simple co-operative nature of the Group had to change into more formal and structured arrangements. A Business Manager and a Management Assistant were appointed in 1993 and 1994, respectively, and in 1995 the formal organisational (and legal) structure for the Group comprising the EuroQol Association and Foundation was inaugurated, with the Business Management office set up in Rotterdam. In addition in 1995 membership of the Group was placed on an individual basis (previously members were primarily associated with specific institutions). Also notable for 1995 was the first use of the term 'EQ-5D' to describe the Group's instrument.

2.3 Biomed EQ-net project 1998-2001

The next noteworthy period in the Group's history was the EQ-net project funded by the Biomed programme of the European Commission (EC). As the decade progressed a considerable number of studies were undertaken by members and were discussed at Plenary meetings, leading in many cases to further work on various aspects of EQ-5D. The EQ-net project provided a timely opportunity to put this work into a more structured context, with most of the efforts of the Group devoted to the project during this 3-year period.

The overall aim was to support the development, adaptation, and application of EQ-5D. After a phase of fragmented multinational research by members of the Group there was an urgent need to integrate and analyse at the European level the data generated in these projects. The tasks involved were divided into 3 sub-projects: Translation, Valuation, and Application. In addition the communication of information and knowledge about EQ-5D was to be addressed. A project management structure was

set up. Notable about this framework is how highly structured it was, with its crafted study programmes and formal timetabling. The Group had discovered that it was not always possible to co-ordinate the various national research projects optimally because of the different national policy and research contexts. Hence each component area had a manager to facilitate the dynamics of the decision-making process, and a series of meetings was scheduled to discuss and solve the issues necessary to attain the project's objectives. In addition the content of Plenary meetings was strongly influenced by the requirement to fulfil these objectives.

The *Annual Activity Report* to the EC issued in March 2001 was the final report on the EQ-net project. Detail on all aspects of the project was subsequently published in book form [2]. Since the main aim of the project was to harmonise data on the valuation of EQ-5D health states collected in different European countries, the main effort was focused on the Valuation sub-project. Two databases were established, one containing visual analogue scale (VAS) valuations and the other time trade-off (TTO) valuations. The Communications sub-project delivered the EQ-net book and both external and internal newsletters. The Application sub-project produced standard operating procedures (SOPs) for the design, analysis and reporting of EQ-5D in clinical, economic and population studies, which were included in the book alongside guidelines for differing modes of administration of EQ-5D: versions for observer, face-to-face administration, proxy, and telephone. These versions were now routinely distributed by the Business Management office. In addition a user conference was held in April 2000 in Amsterdam. The Translation sub-project produced more EQ-5D translations than set out in the original project proposal. Guidelines for translation and adaptation into different languages, together with the definition of EQ-5D concepts were routinely distributed to translating agencies as translation aids.

The statement in the report: "the EQ-net project has provided a powerful stimulus for further development and dissemination of EQ-5D" neatly summed up the significance for the Group of the EQ-net project. Relevant aspects of the EQ-net project are included in later sections of this paper. Full details of the work undertaken are given in the book published as one of the project deliverables [2].

2.4 The period 2002-2009

Following the successful completion of the EQ-net project, and as the resources available to the Group from licensing the use of EQ-5D steadily increased, a more 'programmatic' approach to the Group's activities was implemented. This manifested itself in a number of ways. One was the 'Cheerleader' programme, which was up-and-running in 2002. The other was the establishment of Task Forces in 2005 to focus on particular areas of the Group's work.

Over the period 2002-2005 a series of projects was funded covering a range of topics. It is worth listing these, as some indication of the Group's priorities can be discerned, and this clearly fed into subsequent developments. The topics were: comparison of valuation methods used to generate EQ-5D and SF-6D value sets in the UK; examination of ethnicity, socio-economic status, age and gender in regard to EQ-5D in Cape Town; improving the EQ-5D VAS valuation questionnaire; investigations of the parametric relationship between EQ-5D values elicited with the TTO method and the VAS method; examination of the influence of depression on agreement between proxy assessment and patient self-assessment in stroke using the EQ-5D and SF-6D; EQ-5D related to international differences in self-reported health problems by age, sex and educational level; increasing the levels to five in the EQ-5D descriptive system: application to a youth general survey in Navarra, Spain; and the development of an adapted version of the EQ-5D for infants under the age of 18 months.

Evident here were: a continuing focus on aspects of EQ-5D valuation, EQ-5D in the context of population, demographic and socio-economic variables, the 5-level issue, and the interest in adapting EQ-5D to the child / youth context.

The following Task Forces were instituted: Increased Level (November 2005), Valuation (June 2006), Disease Areas / Condition-Specific (2005), Child-Friendly / Youth (January 2006), Population Health (November 2007), and Digital (May 2008). These generated a considerable amount of activity which had significant impacts on the development of the Group. Specific aspects of the work of these task forces are considered in this paper.

A final important feature of this period was the promulgation of the *Paris Protocol* in 2009. A brief summary is provided here, with a fuller treatment in section 6.8. For many years the Measurement and Valuation of Health (MVH) protocol developed at the University of York in the early 1990s had been used for evaluation purposes (especially economic evaluation). The valuation protocol was raised at valuation task force (VTF) meetings, firstly in 2007 and then in February 2009. It was determined that detailed guidance was required if the Group wished consistent methods to be applied across studies. The protocol was promulgated at the VTF Meeting in Paris in September 2009. Subsequently termed the 'Paris Protocol', it is recommended to all users of EQ-5D-3L but is not mandatory.

2.5 Post-2009 developments

Largely as a consequence of the Task Forces' activities a number of notable developments have taken place in recent years, and these are considered in detail in later sections. The two major developments

were the 5L version EQ-5D-5L, with considerable attention being paid to the valuation of 5L states within a multi-country framework, and the adaptation of EQ-5D for children and young people as EQ-5D-Y. Steps were also taken towards developing 'bolt-ons' to EQ-5D.

Also implemented, in 2013, was a structure of WGs to replace the Task Forces in order to carry forward the activities of the Group. The work of these WGs is included in the relevant sections below. Finally, the WG structure has itself been reorganised in the light of the recommendations emanating from the 2015 *Statement of Research Priorities*.

3. The early years 1987-90

The people who founded what was to become the EuroQol Group met in an environment in which generic (multi-attribute) health status measurement had largely become dominated by the profile approach. Measurement instruments were often lengthy and did not, in the main, provide numbers or indices which would allow the quantitative comparison of composite health states. Although there are considerable benefits for health decision-making in profiling health states, the lack of these indices vitiates against the use of profile instruments in, for example, economic evaluation of health programmes. In addition the variety of available instruments had led to a complex situation with respect to such issues as the number of dimensions that were appropriate for generic instruments.

The first meeting of the Group set out to search for a 'common core' of basic information to be collected by all investigators in the same way. This essential aim came to be crystallised in a set of objectives which can be summarised thus:

- To develop a generic instrument to describe *and* value HRQoL, providing both a descriptive profile and an overall index.
- This instrument was to be a standardised tool to facilitate the collection and pooling of a common data set.
- It was to be self-completed and acceptable for use in postal surveys.

The requirements to meet these objectives, and their implications, were discussed extensively at the early meetings. By the time of the 4th Meeting in July 1988 a detailed protocol had been produced for collecting core data, encompassing 6 objectives, 3 guiding principles, 3 elements concerning descriptions of health states, and 2 aspects of the valuation task. This was accompanied by 2 annexes laying out 6 domains/dimensions and the corresponding items (later termed 'levels') within dimensions, and the 'restricted core' (8 states) and 'extended core' (13 states). These cores comprised

the basic framework envisaged for the collection and pooling of data across Group members. Also put in place was the 111111 nomenclature which in its 5D form has remained in place to this day.

The final version of the 5-dimension EuroQol instrument was developed by 1990 after 7 meetings, letters, phone calls, and position and working papers (but no e-mails!). This intense work took place within a multi-disciplinary group (various medical professions, psychology, economics, sociology, social administration), and across 5 countries, namely the Netherlands, UK, Sweden, Finland, and Norway. English was the 'reference' (later termed 'source' language) with the instrument being developed simultaneously in the other 4 languages.

A key feature was that the work proceeded evidentially. The breadth and depth of the activities undertaken in this process was subsequently captured in Erik Nord's *EuroQolus* classification system for pieces of intelligence generated by Group members [5,6]. His framework comprised 14 design categories: descriptive system, covering letter, preamble, self-rating, thermometer in main task, states to be valued, background data, selection of states, the treatment of unconscious, the treatment of dead, translations, modelling/estimating values, the instrument in general, and miscellaneous. In addition he used 6 response categories: response rates/problems with task, valuations of health states, logical consistency of valuation, reliability of data, interpretation/meaning of valuations, and comparisons with valuations based on other techniques. Very few of the design categories contained no citations. Valuation issues were significant, both in the design and response categories. Other important design categories were the descriptive system, and the treatment of dead.

The ethos of the Group, then, was to develop the instrument on a co-operative basis, proceeding in an iterative way as aspects of questionnaire design and health state valuation were investigated by the constituent centres of the Group. A key feature was that each step of the process was subjected to empirical testing where appropriate with samples of respondents drawn from a variety of sources. The language of the summary notes produced for each of the meetings to the effect that "it was agreed that" is testament to this co-operative approach, with agreement often following the insights gained from the empirical work, and after often very intensive debate.

Many of the issues discussed in these formative years were subsequently brought up again, often years later. Death, duration, and how to model health states, especially in the valuation context, were to be revisited on many occasions. Subsequent sections of this paper treat these issues thematically rather than chronologically.

4. Descriptive System

This section covers the descriptive system in more detail, from its initial construction *via* issues raised in connection with the system over the years through to the development of the 5L version of EQ-5D i.e. EQ-5D-5L. The objectives of the Group outlined earlier led to a number of requirements for the descriptive system:

(i)   The dimensions should be relevant to patients across the spectrum of health care and to members of the general population.

(ii) The descriptive system should be fairly simple - using as few dimensions as possible, with as few levels as possible within each dimension. This would further the aim of generating a feasible number of potential health states for valuation purposes where each health state could be represented by a single index score.

(iii) The health state description needed to be fairly short, and sufficiently clear that the respondent could identify differences between the states.

(iv) The instrument should be amenable to self-completion in a range of settings e.g. in a busy hospital clinic or in a respondent's own home, should be simple enough not to require detailed instructions, and the descriptive page should only take a couple of minutes to complete.

4.1 Choice of dimensions

Dimensions were chosen based on a conceptual process rather than by statistical means such as factor analysis, and were identified through a review of other generic health status measures. This approach was consistent with the original purpose of the Group in trying to 'simplify' the health status measurement world. The dimensions independently suggested for inclusion by group members were broadly similar, with any differences mainly relating to the names of the dimensions rather than their content.

4.2 Number of dimensions

Having originally decided on 6 dimensions, as reported in the initial corporate paper [7], these were reduced to 5 at the 7th Meeting in September 1990 *viz*: *mobility, self-care, usual activities, pain/discomfort*, and *anxiety/depression*. Studies were undertaken with an energy/ tiredness dimension but it was not incorporated into the EuroQol classification, a decision made at the 1991 Lund Plenary meeting.

The next experiment on adding another dimension – *cognition* – was presented at the Rotterdam Plenary Meeting in 1997. It was concluded that the EuroQol classification system could benefit from the addition of a cognitive attribute and also that the introduction of more levels could lead to a refinement of the system. This work was subsequently published [8]. However, at an Executive Committee meeting there was a detailed discussion on the issues raised in adding a dimension to EQ-5D. There was concern that the results of the research reported on adding a cognitive dimension should not create the impression among users that the EuroQol Group as such had investigated adding a sixth dimension to the instrument. **T**his version is not recognised today as an official product of the Group.

On a number of later occasions when the issue of adding dimensions resurfaced it was recommended that there was no need to increase their number. The issue effectively re-surfaced, however, with the concept of 'bolt-ons' (see section 4.7).

4.3 'Standardised' or 'harmonised' EQ-5D versions

This issue was first raised at the 1994 Brunel Plenary. Between April 1994 and the Plenary in September, a decision was taken to remove the valuation part of the standard questionnaire. Until then the EuroQol instrument consisted of what was described as the 'warm-up' task (page 2 descriptive system plus page 3 own health VAS) and the valuation exercise using the VAS approach. The Group had observed that pages 2 and 3 taken together were being used as a self-completed, possibly stand-alone, instrument in clinical studies and trials, so valuation was an option but no longer part of the instrument. In a later paper Kind and Macran observed that what the Group had originally viewed as the warm-up task for respondents to familiarise themselves with the subsequent health status valuation task was shown in a variety of population surveys to be capable of independently capturing information on self-reported health status [9]. "What had started life as a mechanism for investigating the values associated with theoretical health states became a means of collecting information about health status."

At the Hannover Plenary Meeting in 1998 an agreed harmonisation of two EQ-5D English versions that were in circulation, the 'blue' EQ-5D version emanating from York University and the official EQ-5D in Pagemaker was tabled. Improvements to the English language of the EQ-5D were incorporated. This exercise simply effected some changes in wording and minor layout changes, no substantive changes in EQ-5D being made.

4.4 Number of levels

In the early to mid-1990s the issue of the number of levels was a feature at Plenary meetings. Increasing the number of levels to 4 or 5 was initially broached at the 1993 Rotterdam Plenary where one paper [10], plus two research notes on a possible increase to 4 or 5 levels were presented. The meeting minutes stated: "the refinement of levels … was not accepted as standard, though experiments are welcomed".

Next at the 1994 Brunel Plenary Meeting there was comment on the lack of sensitivity between levels 2 and 3. It was stated again that members were free to experiment but 5x3 would remain. Two small scale studies with patients with intractable epilepsy were presented at the Barcelona (1995) and Oslo (1996) Plenaries [11,12]. There was some evidence that this patient group would benefit from the inclusion of an additional level in the descriptive system, indicating a 'slight' or 'moderate' degree of problems.

The issue returned at the York Plenary in 2002, where Kind and Macran investigated extending EQ-5D to 5 levels [9]. The 5L version proved satisfactory in terms of respondents completing the task and their responses were compatible with those obtained from the standard 3L form, and the 3L version was not compromised by the 5L version.

Follow-up came at the Executive Committee in May 2004, when it was proposed that a review paper look at the available evidence with regard to increasing the number of levels. At the Executive meeting in February 2005 it was reported that 3 reviews of a 5L version had been received, which had focused on labelling and on the feasibility of backward compatibility with the 3L value sets. The reviewers were unanimous that there should be labels for the 2 additional levels but were not clear, however, how the labels could be agreed. The 5L version should be based on the existing 5 dimensions of EQ-5D, only 2 levels per dimension to be added. It would be imperative to have valuations for the 5L system but in the short term the possibility of backward compatibility with existing datasets based on 3 levels should be explored. Labelling and valuation were evidently key issues for consideration, hence it was decided to set up a task force to look into labelling choices, a notable landmark in the history of the descriptive system.

The Committee also approved two proposals from the Netherlands for funding related to the 5L system, one on comparing the standard EQ-5D 3L system with a 5L version, and the second on adjustment of EQ-5D TTO value sets for use in an EQ-5D 5-level descriptive system. Papers on these two topics were presented at the next Plenary in Oslo later that year [13,14]. Another paper, from

Spain, considered the valuation of 5L states and back-compatibility with existing 3L valuation sets [15].The Executive Committee meeting in Oslo duly appointed the task force mooted earlier in the year for the increased level descriptive system.

4.5 Increased Level Task Force

The Increased Level Task Force and its offspring the Labelling Sub-Task Force met or held teleconferences 9 times between November 2005 and March 2008. It produced a consensus paper in autumn 2008.

The 1st meeting, held in November 2005, is worth considering in a little detail as the objectives and the potential work of the task force were discussed at length. The meeting had received a position paper which suggested 3 scenarios: (i) Retain the current system but solve the most obvious problems, i.e. modify some of the levels, chiefly within the self-care and mobility dimensions. (ii) Addition of 1 level in each dimension (to produce a 4-level version). (iii) Addition of 2 levels in each dimension (to produce a 5L version).

Reasons for increasing the number of levels were that the current EQ-5D (3L format) was perceived as lacking sensitivity and with having a ceiling (and possibly floor) effect. The number of dimensions was not an issue, but the dimension specification lacked uniformity. In addition there were translation and valuation issues.

During discussion the historical point regarding the descriptive system was reiterated that EQ-5D had originally been developed as a linking tool with other health measures but it had been increasingly used as a 'stand-alone' instrument. The descriptive system was now much more important as its primary use had become as a measure for change in clinical or economic evaluation.

The Task Force produced a wide-ranging set of recommendations for the Executive Committee:

- Agreement to move forward on increased levels to either 4 or 5.
- The current wording of the 3L may require change for some dimensions in the new increased level version.
- Labels should: discriminate more for clinical purposes, demonstrate reliability (test-retest), and be easily translatable into different languages.
- Labels for the levels should be harmonised simultaneously in a number of specified countries covering a wide geographical spread reflecting Group membership.

At the 2nd Meeting in March 2006 a key discussion took place on whether to go for a 4- or 5-level version. Issues raised covered, *inter alia*, sensitivity, burden of translation, number of labelling options, reliability and face validity, number of health states, costs of translations, and ease of use in postal surveys. Following a wide-ranging discussion, and after taking a vote, one of the few occasions that this has occurred in the history of the Group, the Committee decided to recommend the development of a 5L version of EQ-5D.

On labels for additional EQ-5D levels: 5 (labelled) levels was agreed; response scaling was agreed as the general methodology to determine the labels; and a pilot phase – a selection of a set of labels – should decide on whether these labels be based on frequency, capacity, or intensity.

The task force name was changed for the May 2006 and subsequent meetings to the Labelling Sub-Task Force, which facilitated and reviewed a series of funded studies in a variety of locations. Work from these studies was reflected in papers presented at the Barcelona (2006), the Netherlands (2007), and Baveno (2008) Plenary meetings.

Labelling Sub-Task Force Consensus Report: *The development of a five-level version of EQ-5D: progress and process - October 2008.*
This document was a very useful version of the rationale for the Task Force's work and a good summary of the outcomes of the process. It described the progress made with respect to the development of a 5L version, the decisions that were made, the reasons for the decisions, and the remaining decisions required.

The development of a new version of EQ-5D (or any preference-based health status measure) involved 3 main decisions: (1) Dimensions (how many? which?) (2) Levels (how many levels? which labels?) (3) Value set (choice of valuation method, choice of model to analyse data, creation of a value function).

At the 1st Task Force meeting in November 2005 it was decided that there would be no change in the number or nature of the current dimensions for the new expanded version of EQ-5D, as the standard EQ-5D had demonstrated strong measurement properties, had been used to assess HRQoL in a large variety of disease areas, and involved different language versions. A practical benefit of this choice was that the new version should be easy to compare with the standard 3L EQ-5D.

With respect to levels and labels the main choices were: (i) Number of levels for each dimension. (ii) Choice of method for testing the labels. (iii) Criteria for choosing the labels.

On the number of levels, 4 published studies demonstrated that a 5L version of EQ-5D significantly increased reliability, sensitivity (discriminatory power) and feasibility, and broadened the measurement continuum [13,16,17,18].

On labels, the objective of the empirical investigation carried out by the Sub-Task Force was to identify appropriate labels for the expanded version of EQ-5D. 'Appropriate' referred not only to psychometric properties, but also to the property to retain the 'distance' between labels across different translations of those labels. Studies were undertaken in English, Spanish, Chinese, and French. A response scaling approach was chosen to identify appropriate labels. Pilot studies were conducted simultaneously in Spain and the UK, and the main study was conducted simultaneously in France, Spain, and the UK [19,20]. A total of 9-12 labels per dimension were rated. After the consensus report the Chinese study was completed in 2010 and eventually published [21].

With respect to the criteria for choosing the labels, before the labelling studies commenced decisions taken by the Sub-Task Force were: (i) Labels would not have to be similar to the labels of the standard EQ-5D. (ii) The Mobility label would be changed from 'confined to bed' to 'unable to', analogous to the extreme response categories of the other dimensions. (iii) In the Usual Activities dimension in English 'doing' should replace 'performing' to achieve a more colloquial wording. (iv) Level 1 on Self-care was changed to read 'no problems in washing or dressing'.

Criteria were adopted to identify appropriate labels with respect to: position on the valuation space (medians): type of indicator ('qualitative'/ 'quantitative' labels - e.g. 'severe'/ 'many'); commonality; translatability; quantitative and qualitative performance (e.g. variance/ range). At the Labelling Sub-Task Force Meeting in November 2007 labels for a 5L version of EQ-5D in all 3 languages were derived following these criteria. In addition, from the study reports 3 new criteria could be distinguished: (i) preferences/comments of respondents, (ii) colloquial language, and (iii) consistency over dimensions [20,22].

There was now enough evidence to choose final 5L English and Spanish versions, and to harmonise the English, Spanish, and French versions. The next step would be field testing the proposed 5L descriptive system in parallel with the standard version, in various disease states/conditions in order to evaluate the psychometric properties (sensitivity, validity, reliability), of the new 5L measure, and to compare the new measure with the standard version. An important outcome of these studies would be a set of weights that facilitated conversion of 3L preference-based algorithms to the 5L system: a crosswalk between the standard EQ-5D and the 5L versions, based on the current value sets for the

17

standard EQ-5D. This was accomplished and is reported in section 6.8. Finally the Consensus Report referred to valuation studies for the 5L version, also treated in the latter section.

4.6 EQ-5D-5L approval

In November 2008 the Executive Committee approved the English and Spanish 5L versions as official EuroQol products, and the French version was approved shortly afterwards. This was a significant milestone for the Group: after almost 20 years of focus on the 3L EQ-5D, the Group now had a new 5L version of the EuroQol instrument. EQ-5D-5L was officially launched at the EuroQol Group symposium during the ISPOR Meeting in Orlando, USA in May 2009.

A position paper on the whole process was published [23]. In addition the English, French and Spanish versions provided the basis for translation into more than 25 language versions of the EQ-5D-5L, supervised by the Translation Committee. This subsequently became the Version Management Group (VMG), and is now the Version Management Committee (VMC).

4.7 Bolt-ons

On several occasions over the years the issue of adding dimensions to EQ-5D was broached, notably in the case of cognition, as noted in section 4.2, but it was always recommended that there was no need to increase the number. The issue effectively re-surfaced, however, with the concept of 'bolt-ons'.

At the 2005 Oslo Plenary, the Group implemented a product development strategy, one aspect of which concerned the use of EQ-5D in different disease areas, and one of whose purposes would be to investigate if there was potential to develop a disease-specific short tail, to test it and field in a pilot study. These 'tails' were later more commonly referred to as 'bolt-ons'.

In February 2010 the Executive Committee decided to support bolt-on initiatives as part of the Group's research agenda. A meeting on the way forward was held in the same month. Research on bolt-ons to that date was summarised, comprising: 2 cancer studies, skin disorders, visual impairment, hearing impairment, psoriasis, sleep, and cognition. The agenda itself covered: an overview of initiatives, considerations in the development of bolt-ons, scientific issues, and strategic considerations. The reasons for developing bolt-ons were explored: (i) User demand: there had been a perceived lack of content coverage and/or responsiveness of EQ-5D. (ii) Government (especially in

the UK), industry, clinicians and researchers were interested, and funds were being made available. (iii) There had been a trend in the literature towards disease-specific utility measures. (iv) EQ Group endorsement.

In April 2010 bolt-ons were raised for the first time at the VTF, which had an interest in the valuation aspects, would like to have an input in the way these valuation issues might be addressed, and would work with the CSPTF (Condition-Specific Task Force) in this area. Later in the same year the Athens Plenary (2010) included a pilot bolt-on study testing a version of EQ-5D-5L specifically for use with people with psoriasis, EQ-5D-Psoriasis [24].

The design and valuation of bolt-ons has largely proceeded within the same structural framework as the basic EQ-5D approach by having, say, for each bolt-on in a study, 3 three severity levels to match the EQ-5D-3L. Resulting health states are then valued by, e.g., a sample of members of a country's general public using the traditional EQ-5D TTO approaches.

A bolt-on working group under the remit of the CSPTF was appointed in September 2011. At that month's Oxford Plenary there was a session devoted to bolt-ons: a paper on cognition as a bolt-on dimension [25],and another paper on 3 'add-ons' (hearing, vision, tiredness) to EQ-5D [26].

Meanwhile the YTF had been working on developing a cognitive dimension as a bolt-on for the EQ-5D-Y in Germany, and a paper was presented at Oxford. The dimensions comprised: *cognitive abilities* (working with the mind), *concentration* (for example, in school, while learning, while reading, while doing sports), *retentiveness* (memorising, for example, things learned at school, homework, intake of tablets, appointments), *school performance* (for example, reading, maths, learning, homework, exams).

A paper on estimating index values for EQ-5D-Psoriasis was presented at the Netherlands Plenary [27]. Cognition as a bolt-on was revisited in the EQ-5D-5L context at the Montreal Plenary [28]. Although according to the latest research strategy commencing in 2015 bolt-on studies have not been designated as an explicit research priority, following detailed discussion at the Krakow Plenary meeting in September 2015 the Descriptive Systems and Valuation Working Groups are expected to re-visit bolt-ons.

4.8 Concluding remarks on the descriptive system

This section has covered the entire period of the descriptive system, from the early formative days which resulted in a 'settled' system which remained unchanged for many years, except for minor amendments. This descriptive system is still with us, of course, embedded in what is now called EQ-5D-3L. The development of EQ-5D-5L followed over 20 years' concentration on EQ-5D-3L. Once this was accomplished, 5L work moved on to valuation issues, and comparisons of EQ-5D-3L and EQ-5D-5L in different contexts were well underway. These features were evident from papers presented at Plenary meetings from 2010 onwards. The 3L, in which so much time, effort, and resources were invested, can be expected to continue to play a part in the Group's activities alongside the investment now being made in the 5L: currently (June 2015) there are 126 language versions of EQ-5D-5L.

5. Translation, language and semantics issues

At the outset of the EuroQol enterprise English was used as the working language: Group documentation over the years also refers to it as the 'source', 'reference' or 'standard' language. Those members of the Group whose first languages were not English helped to construct the EuroQol instrument in ways that the language and concepts being used and the constructs being developed using English could then be applied in the context of their own languages. Issues of semantics, wording and language were well to the fore from the early days. For convenience we can subsume these issues under the portmanteau term 'translation'.

5.1 Translation issues 1992-98

A valuable source in tracing the history of translation issues is Herdman *et al* who noted that various recommendations and guidelines had been produced within the HRQoL field which aimed to secure equivalence between different language versions of an instrument [29]. Two overall approaches were possible: (i) forward and back translation complemented by panel discussion and patient testing; (ii) simultaneous development of an instrument in various national settings, thus permitting an exploration of the extent to which elements or facets of HRQoL were common to the various countries involved in its development. Early versions of the EuroQol instrument followed the second approach, being developed by researchers from 5 European countries - Finland, the Netherlands, Norway, Sweden, and the UK: "in a process which allowed researchers from all those countries to contribute to the instrument's content and design."

Herdman *et al* stated: "However, as demand for the instrument grew in countries where no language version ... existed, it became necessary to develop and implement translation procedures which would ensure a certain degree of rigour when translating the instrument for use outside the original group of languages." This work was initially stimulated in 1992 when researchers from Spain became involved in the Group, with the instrument being translated and validated in Spanish and Catalan. This Spanish version was produced by forward and backward translation, followed by testing the translated versions on a panel of members from the target population; thus following the first approach noted above. In this version respondents were asked for the first time in the EuroQol context to provide ratings of clarity, use of colloquial language, and the appropriateness of the questionnaire content, all rated on a 1-10 scale [30].

Evidently the Group was aware of the necessity to standardise the methods and procedures used to produce translated versions of the instrument. The procedures used for the Spanish and Catalan versions served as a model for draft translation guidelines developed between 1994 and 1995. The essential features (listed in [29]) were:

    (i)   Production of 2 (non-literal) forward translations by 2 translators working independently.

    (ii)   Forward translators should be native speakers of the target language but fluent in English, and at least one of the forward translators should be a professional translator.

    (iii)   Production of a first consensus version was based on combining the 2 forward translations in a meeting between translators and the research team.

    (iv)   Production of 2 back translations from the first consensus version.

    (v)   Back translations produced by 2 professional translators working independently, although in this case the translators should be native English speakers fluent in the target language.

    (vi)   Cognitive debriefing in the form of testing the final version of the questionnaire in a small number of potential respondents to determine clarity, acceptability, and relevance.

The next language versions along after these guidelines were promulgated were the French and German versions derived from the English version.

Translation issues continued to exercise the Group, so that at the 1995 Barcelona Plenary a paper contemplated challenges for the future with respect to international language versions [31]. The main conclusion was: "Having reviewed the progress of translation over the past 9 years within the EuroQol Group, we believe the most significant gap in the Group's research regarding translation to be the lack of an attempt to directly assess conceptual equivalence between the English version and each of the other language versions." In addition, the Group should consider the wording of the English version in relation to criticisms raised from the translation process. The minutes of the Business Meeting in Barcelona included a lengthy section on these issues. For EQ-5D it was agreed

that it was essential to have a consistent set of questionnaires in a standardised format in the different language versions. With respect to the English source version, some English-speaking members of the Group considered that the language of the source version needed improvement. There was concern, however, that any changes might affect the validity of the research already carried out.

At the Oslo Plenary in 1996 a paper on the semantic equivalence of EQ-5D delivered 3 major findings [32]: (i) A whole range of implicit meanings was highlighted which would be helpful in translating EQ-5D into other languages. (ii) It was clear that there were particular words/phrases with which the Group was in agreement. (iii) Several alternative words/phrases had been suggested which should help the process of choosing words in alternative languages. The Group should take decisions in 3 areas: (i) On specific issues of semantics such as extending the range of alternative words and phrases, and agreeing the meaning of selected terms and phrases. (ii) To consider general issues, for example, to decide on the level of importance the Group wished to give to the issue of translation, and to agree on a protocol clearly linked to particular notions of 'equivalence'. (iii) To decide whether the Group was interested in pursuing the use of EQ-5D as a tool to investigate issues in the translation of HRQoL instruments *per se.*

The issue of translation guidelines was also raised in Oslo. A number of members offered to consider the state of the art regarding guidelines and if necessary develop proposals for discussion by the Group: this was the genesis of what became the Translation Committee. At the March 1997 Executive Committee it was reported that the existing guidelines were being expanded. At the end of the translation process for each language, the Translation Committee would make recommendations to the Executive Committee regarding the eligibility of the various translations to receive "an official EuroQol stamp of approval".

5.2 Biomed EQ-net project

'Translation' was one of the arms of the project. This term was viewed broadly enough to encompass issues concerning the nature and meaning of health and HRQoL in the context of EQ-5D, its construction, design, and wording. As noted earlier, when EQ-5D spread to languages other than the original ones in which it was developed, the process of translation began to point to difficulties in language usage, and more fundamentally, to differences in the conceptualisation of entities such as EQ-5D dimensions and items, across countries and languages. This experience, amongst other considerations, led the Group to consider more closely the meanings of concepts and the related wording used in EQ-5D, not least in English. The EQ-net project enabled resources to be allocated to these issues and Chapters 10-12 of the project book reported in detail on this [2]. There was also

additional relevant material in Chapter 13 of the book which reviewed work on EQ-5D in a number on non-European countries.

Chapter 10 concerned the interpretation of EQ-5D concepts [33]. It considered 3 main areas: the range of shared and different meanings among EuroQol Group members, interpretations of EQ-5D amongst people who had completed the questionnaire in UK English within the UK, and interpretations amongst people who had completed the EQ-5D in other languages. There were evident variations in interpretation of EQ-5D concepts by Group members themselves. This may appear surprising given that work on the instrument had proceeded since 1987. One reason could be that members were required to try to be explicit about concepts and terms which often contained implicit meanings. Other reasons could be: the view of a group was likely to change over time, asking individuals to think of the group view was likely to produce variety, group members would hold different views individually which reflected personal experiences of life, and an instrument developed by a group would represent a compromise of opinions.

The main recommendations for further research included: (i) broadening the range of methods used to investigate meaning; (ii) broadening the base of investigations to include a wider range of countries and disease groups; (iii) using quantitative and qualitative measures alongside each other to investigate the relationship between interpretations of words and scales and valuations of health, in relation to socio-economic, demographic, and geographic variables.

An appendix in the book provided a definition of EQ-5D concepts which was influenced by the research on which Chapter 10 was based. Moreover, a number of the interpretations provided by the research exercise came to be included as part of the translation package sent to researchers undertaking new versions of EQ-5D [29].

Chapter 11 reviewed work undertaken on producing other language versions of the EQ-5D. A brief history of translating the instrument was presented, followed by a discussion of translation SOPs and the 'quality control' of the translation process [29]. The conclusions of the chapter were that: (i) development of new language versions of the EQ-5D reflected developments in the area of cross-cultural adaptation and in the HRQoL field in general; (ii) local researchers should be aware of the need to test a new version's measurement properties of reliability, validity, and sensitivity to change; (iii) analysis of similarities and differences between values obtained for EQ-5D health states in different cultural settings may be used, alongside careful qualitative research, to determine whether there were differences in values between different respondents in different countries, and as a further check on the extent to which meaning had been transmitted successfully between languages.

Another appendix in the EQ-net book contained a taxonomy of definitions of EQ-5D concepts. This was made available to researchers wishing to develop new language versions, alongside the SOPs, which were reproduced in further appendices. One contained detailed translation guidelines, ordered under the following heads: forward translation, back translation, respondent testing, forward translations into the target language, production of first consensus version, report on the forward translation process, back translations of the consensus version into English, back translation meeting, report on the back translation process, and finally a report on respondent testing. The other appendix detailed the interview process and a series of questions concerning the respondent's experience in using EQ-5D, ranging from overall issues such as ease of understanding, length of questionnaire, and clarity of instructions, to more detailed probing of meanings of words and phrases.

Chapter 12 contained a detailed account of an exploration of the results of translating the EQ-5D into 11 European languages: Croatian, Czech, Danish, Dutch, Finnish, French, German, Italian, Polish, Portuguese, and Spanish [34]. Information on the translation process in the HRQoL field had rarely been written up in the international field of health outcomes assessment, so the material presented in this chapter was offered as an insight into the translation process, in particular the difficulties encountered and how they were resolved, and how health state valuations might be affected. First, there was a fascinating description of the translation of the EQ-5D dimensions into languages other than English, replete with interesting examples. Then a number of studies of the potential impacts of decisions made during the translation process on valuations were presented.

A number of pointers to future work emerged from this preliminary, and essentially qualitative, exploration: (i) The need to explore and achieve good quality translations that were semantically equivalent. (ii) Assessing the impact of differences in the translation of questionnaires had to be supported with empirical research. (iii) The acceptability and conceptual equivalence of existing translations could be further evaluated using both qualitative and quantitative methods. (iv) For reliable international comparisons additional aspects of conceptual, item, operational, measurement and functional equivalence needed to be investigated
.
Chapter 13 focused mainly on valuations in selected countries around the world: Japan, New Zealand, the United States, Canada, and Zimbabwe [35]. There was some commentary from the different countries on translation issues. The Japanese section contained specific examples of how words and phrases in the English EQ-5D were translated into Japanese and showed the cognitive/conceptual problems that could arise. The Zimbabwean section discussed translation and conceptual issues with regard to one of the major language groups, Shona. The authors concluded that, despite Shona people having a different view on aspects of health status from that portrayed in the EQ-5D, it would still be possible for people to respond to the questionnaire, and this was borne out in the empirical results.

The New Zealand section examined health valuations across the 3 major ethnic elements of New Zealand society, namely Maori, European/Paheka, and Pacific Islands people. Despite the possibility of different conceptualisations of health, these did not appear to be borne out by the valuations.

5.3 Translation issues after the EQ-net project

Two major developments with significant translation implications can be discerned from later years. One was the creation of new EQ-5D 'products', with specific reference to 'child-friendly' versions of the instrument. Following the paper by Hennessy and Kind [36], considerable resources and research efforts were put into child-friendly versions in a number of languages, overseen by the YTF, see section 8.3. This EQ-5D variation is termed the EQ-5D-Y and is available in various languages. From the point of view of the current section, the development of EQ-5D-Y is of considerable interest in that it provided yet another focus on matters of conceptual significance for the EQ-5D, and the issues raised for translation in the child-friendly context.

The other development was the EQ-5D-5L. Translation issues were to the fore in the development of appropriate labels for the 5L instrument, with research initially being conducted in English, Spanish, French, and Chinese.

With the detailed SOPs in place, EQ-5D was translated into a large number of new languages in the new millennium. Also accomplished were language adaptations, such as Spanish for Latin American countries. An appendix in the EQ-net book contained guidelines for such adaptations [2].

5.4 Version Management Group

In November 2009 the Executive Committee decided that a small English language reference group would be formed to advise on version management, e.g. the harmonisation of instruction texts. A report from the VMG presented at the September 2011 Executive Committee meeting is a very useful summary of the activities of the VMG. It had been very active reviewing new language versions as well as responding to client and translation agency queries, updating essential documentation, and implementing systems aimed at improving version control and management. In outline:

(i)   A large number of new language versions for a range of clients and over several platforms (tablet, web, personal digital assistant - PDA) had been monitored and reviewed.

(ii)   Version control, involving improved methods for archiving and tracking versions, and quality control, where errors or sub-optimal wording in existing versions were corrected or improved, was outlined.

(iii)   The VMG commissioned a back translation and review of the original 5 language versions of the EQ-5D-Y from a translation agency.

(iv)   To formalise the approval process for new language versions, a new evaluation form was developed, to be completed on completion of the translation process, which provided a formal sign-off document for each translation process.

(v)   The new evaluation form was applied to formally sign-off on 21 language versions of the 5L. Including the original languages in which the 5L was developed, almost 60 language versions or adaptations would shortly be available.

(vi)   There was a need to update and rationalise adaptation guidelines to take into account in particular the availability of the 5L version and the need for partial translations (e.g. separate instructions for web, tablet or PDA versions). The VMG had collaborated with the Digital Task Force (DTF) to produce templates of 3L and 5L tablet and PDA formats.

5.5 Concluding commentary on translation

This section has traced translation issues from the time of the establishment of the original EuroQol instrument. As the number of language versions increased it became clear that linguistic and semantic considerations needed to be addressed, not least in the English version. The work undertaken for the EQ-net project was crucial in this regard, and the experience gained during the project helped subsequently in the labelling and translation of the 5L and Youth versions. In addition the advent of digital versions of EQ-5D raised its own questions about the design and content of EQ-5D products. Currently what is now designated as the VMC has a key role in the ongoing process of product development.

It can also be noted that the Group has introduced a new system to store and maintain its products ('M-files'). This covers quality assurance, a new layout for EQ-5D products, and a system of naming conventions for these products. Full details are available on the Group website.

In conclusion, the position in 2015 with respect to translations, detailed on the EuroQol website (www.euroqol.org), is 174 EQ-5D-3L and 126 EQ-5D-5L language versions. In addition alternative formats that have been made available comprise: proxy, telephone interview, IVR (interactive voice response) *via* telephone, web, tablet, and PDA.

6. Valuation

6.1 Introduction

This section surveys how the Group has handled the valuation of health status. After describing the early work dealing with valuation issues, the Group's state-of-play paper is briefly annotated to give context to this work and the follow-up efforts by the Group in the early 1990s [37]. One important body of work in this period was conducted at the University of York through the Measurement and Valuation of Health (MVH) project which produced 'tariffs' for EQ-5D states and this project is outlined. The next significant development was the Biomed EQ-net project of 1998-2001, in which valuation was one of the key components, hence this body of work is summarised. The subsequent sub-section simply lists the topics which were the subject of papers presented at Plenary meetings from 2001 to 2007 to demonstrate the variety and breadth of ongoing valuation studies, and this is followed by a brief look at valuation issues in the same period.

As the Group expanded its membership an even greater volume of effort was put into valuation. For convenience this body of work has been subsumed under the heading *Developments after 2008.* In 2009 a formal written-up guide for users of EQ-5D-3L was produced in the form of the 'Paris Protocol'. The motivations for this were essentially twofold: demand for guidance on utilising EQ-5D-3L from potential instrument users, and the desire from the Group for consistent methods to be applied across the burgeoning number of studies being undertaken. Some detail is presented on the protocol.

The next important task was to value the newly-minted EQ-5D-5L, whose construction was depicted in section 4.5. Much of the subsequent (developmental) valuation work by the Group since 2009 has been with the 5L. First, EQ-5D field testing and development of a crosswalk between 3L and 5L is described, then the multi-country EQ-5D-5L valuation programme is examined in some detail. The rest of the section is taken up with a number of matters. The role of the Valuation Methodology Working Group (VMWG) is examined, the PRET (Preparation for the Re-valuation of the EQ-5D Tariff) and FEDEV (Further Exploration of Discrete Choice Experiments with Duration for EQ-5D Valuation) projects in the UK are described, and a review of EQ-5D valuation studies with its recommendation for a checklist for reporting EQ-5D valuation studies is outlined. One of the consequences of the latter work was the establishment of the Value Sets Working Group, whose objectives are listed. Valuation activities are to be co-ordinated from 2015 by a newly constituted Valuation WG. Finally some concluding remarks are assayed on what has been a major component of the Group's activities.

6.2 Early work on valuation

The valuation of health (or more broadly HRQoL) states has always been fundamental to the EuroQol enterprise. The 1st Meeting itself focussed on the data requirements for the basic valuation task: (a) descriptions, (b) time context, and (c) valuation method. The meeting decided that the latter should initially be a rank ordering, followed by an exercise in magnitude estimation (ME), with 0 and 1 as the postulated values of the worst and best states respectively. ME was replaced by a visual analogue scale (VAS) at the 3rd Meeting, and by the 5th Meeting VAS was the method of choice. However, it was suggested that it would be useful if Group members could also use the same descriptors that had been constructed to elicit valuations by other methods.

Considerable emphasis was thus placed in the development years of the EuroQol instrument on the use of visual analogue scales, with page 3 requesting respondents to indicate 'your own health state today' on a VAS 'thermometer'. The subsequent valuation task also asked respondents to place composite health states (11111 and so on) on a VAS. The main reason for adopting VAS as the Group's standard method "was that self-completion questionnaires were seen as the only practical means of obtaining large valuation data sets and the VAS was the most suited to such a survey instrument"[38].

EuroQolus citations show, however, that experimentation and research was also undertaken in the early days with: TTO (Brunel, York, Rotterdam, Lund, Oslo); paired comparisons (York, Oslo); Likert (York), and equivalence of numbers (Rotterdam) [5,6]. In addition there was some consideration of the quality-adjusted life-year (QALY) as a measure of social valuation of health for use in economic evaluation (Rotterdam, Oslo), and of the healthy-year equivalent (HYE) as an outcome measure (Rotterdam). More generally, the importance of valuation was indicated by the *EuroQolus* data, which showed a significant number of citations for valuation issues.

With the later adoption of the TTO approach to derive 'tariffs' for EQ-5D some perspective can be placed on the emphasis on VAS and TTO by the number of citations in the valuation chapter in Brooks [1]: TTO (182), VAS (149), discrete choice experiments (DCE) (113), standard gamble (SG) (14), person trade-off (PTO) / equivalence (14), paired comparisons (4), Thurstone scaling (2), and ME (2).

The first EuroQol corporate paper stated: "The single most significant property which was designed into the instrument by the Euroqol Group was a capacity to yield a single index value for any given health state." The paper also collated results from 3 studies conducted in Bergen op Zoom (the Netherlands), York (UK), and Lund (Sweden) both for the ordinal data from each study, and "if the data are treated as cardinal values, then a regression analysis can be conducted." [7]. It was concluded that whether treated as ordinal or cardinal data, the results of the three studies were "strikingly similar." This greatly encouraged the Group in pursuing its valuation objectives.

In developing an instrument that could provide appropriate health status scaling, it should also be noted that Group members from these early days spent considerable time and effort on two important aspects of valuation, namely the duration of health states and the valuation of 'dead'. Again, this is evidenced in the frequency of citation of items relating to these topics in *EuroQolus*. In addition Group members were conscious of the requirements that the instrument should be reliable and meet (psychometric) tests of validity. A further consideration was the question of whose values (experts, patients, the general public) should be used.

6.3 1996 State-of-Play paper

Most of these issues were treated in detail in the next corporate paper published by the Group which covered activities until the mid-1990s [37]. With reference to the valuation task the paper stated that "the EuroQol instrument is intended to be a generic measure with interval properties that allows for the calculation of QALYs in economic evaluation." It had become clear from studies conducted that VAS scores were not meant by respondents to express trade-offs between longevity and the numbers of people helped, so that VAS scores could not be used directly as weights for life years in QALY calculations. Hence the Group was exploring the use of alternative valuation methods. More generally issues surrounding the valuation task, the appropriate number of health states for this task, and modelling, were treated in the paper.

With respect to the valuation task there had been considerable experimentation, as the EuroQol instrument was constructed, on formats: use of horizontal rather than vertical thermometers; presenting only one health state (i.e. one box) per page; the use of split scales; providing examples to respondents prior to the main valuation task; and verbal explanations of the approach (a compromise between interview-style and self-completion). Some of these design issues were to resurface in later years, in particular with the onset of digital technology.

With the 5-dimension 3L instrument in place from 1990 this gave 243 possible health states plus 'dead' and 'unconscious.' Initial experimentation had indicated that it would be feasible for respondents to value 12-16 of these states. As the paper notes: "The rank ordering of the common core of 13 states using the VAS technique in a variety of settings and with a variety of respondents proved sufficiently robust to suggest that the choice of core states was covering a broad range of different aspects of health status measurement." The solution to the vexatious question concerning how to treat being dead was that 'dead' was separated from the other valuation tasks, but respondents were asked to site it on the same VAS scale as other states so that they could indicate which states were rated better or worse than dead. Studies had also analysed additional health states; one broad approach that was applied was modelling.

The basic aims of modelling for the Group were essentially three-fold: (1) data reduction, e.g. to measures of central tendency such as means and medians; (2) data prediction, e.g. from the 13 core values on which the Group was focussing to the 243 health state values; and (3) data transformation, e.g. from VAS values to TTO values. Much of the work until this juncture had involved data reduction. Data transformation studies were also in progress. In addition work was being undertaken on more formal (statistical) modelling in the context of estimating the values to be attached to the 'non-core' health states. Subsequent to the early work countries represented in the Group have sought to provide national values/tariffs for all 243 states to meet basic aims such as the ability to conduct evaluative studies of alternative health programmes [39].

Finally, on the issue of alternative scaling (valuation) approaches, the paper noted that the VAS had proved simple and appropriate for use in postal questionnaires. However, the evidence was that VAS did not provide cardinal values or utilities, so it was necessary to explore alternatives. A number of studies were reported in the paper comparing VAS valuations with PTO, SG, TTO, and the Rosser-Kind index.

6.4 Measurement and Valuation of Health (MVH)

The studies undertaken as part of the MVH programme at the University of York in the early 1990s were important in their use of the TTO approach to generate values for EQ-5D health states. Since the approach led, in principle, to social valuations, this would make these valuations suitable for use in, especially, economic appraisal. Unlike the VAS method, these valuations would be applicable in the QALY context. This significant development brought forth the terminology of 'tariffs' that could be applied in the evaluation of alternative health programmes [40]. A helpful article which places the

MVH tariffs in the context of what is termed TTO-based 'scoring algorithms' is that of Feng Xie [41]. The MVH tariffs are still in use today in the EQ-5D-3L context.

6.5 Biomed EQ-net project

Construction of a common EQ-5D value set based on VAS values had been a key aim from the outset for the original members of the Group. The EQ-net project proved to be essential in achieving this aim as the Group was able to build up databases for the VAS and TTO valuation approaches, using material from a variety of studies across a number of countries. These studies were not completely standardised, so a lot of effort had to be put into pooling the data in appropriate ways.

Construction of the EQ-net VAS and TTO databases

The work undertaken concentrated on a 'standard' set of 18 EQ-5D health states on which much of the valuation work had been undertaken. The set comprised: the 13 states employed from the outset of the Group's empirical work (the 'common core'), 'dead' valued twice, 'unconscious', and 11111 and 33333 (core states) both valued twice. The VAS database consisted of valuations from 11 studies conducted in 6 countries (Spain, the Netherlands, Finland, Germany, Sweden, and the UK). The TTO database comprised valuations from Germany, Spain, and the UK.

The major points to emerge from this work were: (i) Analysis of respondent characteristics showed some differences between the study populations, which were partly influenced by the differences in sample features, e.g. differences in the methods of collecting valuations. (ii) The number of exclusions and inconsistencies differed significantly between the studies, these differences being connected to differences in the VAS valuation methods used.

Comparison of EQ-5D VAS valuations

The data from the European VAS-based studies conducted since 1991 was pooled in order to compare valuations across studies and countries in a rigorous way by accounting for measurable differences in methods used, study features, and respondent characteristics [42]. There appeared to be a considerable degree of agreement between health state valuations in studies from several European countries, and the evidence analysed suggested that the Western industrialised countries, at least, may share a similar value climate.

A European EQ-5D VAS valuation set

These conclusions led naturally to the question whether a European value set based on VAS scores could be constructed from the database [43]. The aim was to see if the joint data sets of the VAS

scores in the database could be described by one model. Since the database provided data on a range of health states this wider information could be used to model the data in such a way as to assign a unique HRQoL value to each possible health state as defined by EQ-5D (243 in all). The modelling analysis concluded there was strong evidence that it was possible to describe the joint VAS data set by one common model.

It was recommended that the results of this exercise be compared with valuation data from future work with a uniform European, or even world-wide, approach. The results suggested that valuation studies need not be confined to national borders. This was evidently a major landmark for the Group's members, especially those who had been diligently pursuing the original aims set out in those early meetings over a decade before.

Comparison of EQ-5D TTO values

Busschbach *et al* focussed on EQ-5D valuations derived using the TTO approach, and codified in the TTO database, which contained valuations from studies undertaken in the UK, Germany, and Spain [44]. Comparison of these values gave the opportunity to test differences in TTO values between these countries. The practical implications of the analysis were threefold. (i) Since there might be national differences in TTO values, choice of a particular TTO value set, say for economic evaluations, would need to anticipate these national differences. (ii) Within international trials, it might be appropriate to use one TTO-based value set. Given its design and construction and the large sample employed, the MVH value set had been the most often used in this context. (iii) Even where one value set was used in an international trial, it was recommended that the best local valuation set be used for (local) reimbursement decisions.

At this juncture it was evident that the TTO evidence was somewhat more limited than the VAS material. More studies could be expected to be reported, but in general the Group was keen that further work be undertaken using choice-based methods such as the TTO.

Hence by completion of the EQ-net project the position with respect to valuation was that a common value set had been created using VAS valuations, and that the TTO approach as a choice-based method needed further study.

6.6 Valuation-related papers at Plenary Meetings 2001-07

Following the Biomed project, Group members pursued a variety of topics with respect to the valuation of EQ-5D. (1) Continuing work with VAS. (2) Comparisons of VAS and TTO. (3)

Increased work on TTO. (4) Comparison with other instruments, e.g. SF-6D, 15D, HUI2, HUI3. (5) Modelling. (6) 5L. (7) Economics papers. (8) Other topics including: retrospective performance of EQ-5D, valuation of all 245 states, patient valuations, religious values, computer-assisted data collection, Thurstone scaling, PTO, and an international valuation set.

6.7 Valuation Issues 2001-07

Evidently a wide range of issues was being pursued by Group members. Valuation studies were placed on a more Group-orientated basis when during the early part of this period a variety of Cheerleader valuation projects was approved as the allocation of funding increased.

A key decision was also made to inaugurate a number of task forces, one of which was the VTF, whose first meeting in June 2006 is worth a short review as it provides the opportunity to consider the state-of-play with respect to valuation at that time. A number of position papers on key issues were tabled and discussed. With respect to the choice of scaling method, the focus had been on VAS and TTO. There were some mentions for SG, Thurstone scaling, and paired comparisons, and the DCE approach was flagged up and discussed. Risk and time preference needed to be addressed and the ever-present issues of death valuation and duration were raised. QALYs were deemed important especially for economic evaluation, but there were many other applications. Also tabled were user requirements and eliciting patient valuations.

The question was posed: was there consensus about prioritisation in the EQ Group? Points made included: investigating contingent valuation (discrete choice) type work, carrying out more research into economic aspects, the appropriate way to use EQ-5D, the relative resource implications of the alternative valuation approaches, the possibility of combining methods, the desirability of addressing specific research questions, and defining empirical work that could help with valuation choices.

It was decided with respect to priorities for research groups wishing to carry out valuation studies in the immediate future that VAS should be included, and the York MVH protocol be improved, as the perceived weaknesses of TTO had to be addressed.

The March 2007 VTF was notable for being devoted to D*CE* vignettes, thus marking the willingness of the Group to engage with the DCE approach, as suggested at the earlier VTF meeting.

6.8 Developments after 2008

The years from 2008 to the present have been a very busy period with respect to EQ-5D valuation matters. Issues tackled have included the continuing role of the VAS, valuation protocols for 3L and 5L, variations on the TTO approach and DCE methods for the 5L, modelling, medians, the hardy perennials duration and dead, electronic products, the interaction between task forces on potential youth valuations, bolt-ons, and crosswalks.

The new significance of the DCE approach was confirmed at the Board in March 2008, when additional funding for DCE work was provided. The study comprised a head-to-head comparison between TTO and DCE and involved collaboration between a substantial number of study groups.

Valuation Protocol

For many years the MVH protocol developed at the University of York in the early 1990s had been used for evaluation purposes (especially economic evaluation). The valuation protocol was raised at VTF meetings, first in 2007 and then in February 2009. There was concern in particular that the Group had never documented its state of play *re* valuation. A detailed manual was required for people wishing to undertake valuation if the Group wanted consistent methods to be applied.

The protocol was promulgated at the VTF Meeting in Paris in September 2009. Subsequently termed the 'Paris Protocol', it is recommended to all users of EQ-5D-3L but is not mandatory. Researchers wishing to carry out valuation studies should be encouraged to follow the protocol and be asked to contribute to the Group's valuation database.

The modified MVH protocol was based on a document prepared by Paul Kind, who was part of the original MVH team. A total of 10 modifications to the MVH approach were provided, 2 of which were optional. These are shown in Table 1.

Table 1: The EQ-5D-3L Paris Protocol

Paris Protocol
Adopt a fixed block design in all valuation studies in which a partial set of the 243 EQ-5D states is used.
Remove 'unconscious' from the set of EQ-5D health states used in valuation studies.
Amend the procedures for recording valuation data so as to capture the order in which states were presented.
Substitute the word 'dead' for 'immediate death'.
Streamline the VAS rating exercise by removing the fractionation exercise.
Amend the protocol so that respondents are presented with health states for VAS rating in a random order.
Data coding forms for the TTO procedure should include scope for the interviewer to indicate when (if at all) respondents deviate from the verbal instructions they have been given.
In any new valuation study, if resources permit, then the investigation of variable time horizons based on respondent life expectancy would be a useful addition assuming that the study protocol also provides for the collection of utility weights based on a 10-year, fixed time horizon.
Optional modifications
Extend interview procedure to incorporate within-respondent retest and/or to provide internal evidence of the interpretation.
Determine the measure of central tendency that best fits the valuation data and the information needs of those who are expected to use it.

For a number of issues no formal modifications were proposed. Each issue was discussed in the protocol and commentary and general advice supplied. These covered states worse than dead, selection of states, data fallibility, transformation (to utilities), and design of sampling frame. The Paris Protocol thus provided a guide for valuation studies that could be recommended to prospective EQ-5D-3L users. A number of issues were either not suitable for 'hard-and-fast' rules, or still subject to research and deliberation within the Group. Some of these had been long-standing, such as dealing with states worse than dead and transformation. Ongoing work on the 5L version of EQ-5D was also handling these questions.

EQ-5D 5-level Field Testing and Development of a Crosswalk

Having developed the EQ-5D-5L, as approved by the Executive Committee in November 2008, and described in section 4, the Group embarked on the task of valuing the 5L's constituent health states. This was initially accomplished by undertaking crosswalk studies.

In early 2009, the CSPTF began to focus on the development of a protocol that would field test EQ-5D-5L, the main aim being the development of a crosswalk between existing 3L value sets and the new 5L descriptive system. A secondary task was to examine the validity and sensitivity of the new 5L version in different disease states/conditions. The CSPTF developed a protocol that emphasised heterogeneity by proposing to collect data on several different diseases in patients whose health problems spanned the range of levels captured by the new 5L system.

The Group subsequently coordinated a study that administered both the 3L and 5L versions of EQ-5D, resulting in crosswalk value sets for the EQ-5D-5L. The crosswalk was based on a response mapping approach that estimated the relationship between responses to the EQ-5D-3L and EQ-5D-5L descriptive systems, and subsequently established a link to the 3L value sets.

A total of 3691 respondents completed both the 3L and 5L across 6 countries: Denmark, England, Italy, the Netherlands, Poland, and Scotland. Different sub-groups were targeted, and in most countries a screening protocol was implemented to ensure that a broad spectrum of levels of health would be captured across the dimensions of EQ-5D for both the 5L and 3L descriptive systems [45].

As a consequence of this work the EuroQol website contains crosswalk value sets for the EQ-5D-5L for the following countries: Denmark, France, Germany, Japan, the Netherlands, Spain, Thailand, UK, US and Zimbabwe. It also includes a section on the crosswalk model and its methodology, and a Crosswalk Index Value Calculator.

Multi-country EQ-5D-5L valuation programme

The crosswalk produced valuations mainly based on TTO methods. The Group then decided to test the performance of different variants of TTO but in addition took the opportunity to test the potential of discrete choice methods, in both cases within a multi-national study framework.

Context

The requirement to value the 5L instrument was the subject of considerable deliberation over a number of years. A document *Protocol to value the EQ-5D-5L* was produced, and it was noted that the Group had also actively pursued the development of digital technology to conduct valuation interviews, under the auspices of the Development Electronic Valuation Technology (DEVT) project, launched in November 2009. The project entailed the development and testing of a digital 'prop' for the valuation task, replacing the previous reliance on physical props (such as TTO boards and paper questionnaires) and the provision of a structure for facilitating the logistics of 5L valuation studies. The Group then embarked on a multi-country programme which is now considered in more detail to

demonstrate how the Group tackled the EQ-5D-5L valuation task, based on the paper by Oppe *et al* [46].

Introduction

EQ-5D-5L retains the original five dimensions, but has a five level classification of severity - no problems, slight problems, moderate problems, severe problems and extreme problems, and describes 3,125 (5^5) unique states. An interim solution to value these states was accomplished in the crosswalk studies (see above) where a mapping algorithm between the two instruments allowed EQ-5D-5L states to be assigned values from existing EQ-5D-3L value sets.

Value sets for EQ-5D-3L were mainly based on time trade-off (TTO) methods [39]. However, TTO is a complex valuation technique. What can be termed 'conventional TTO' uses conceptually different approaches to the valuation of states better than dead (BTD) and worse than dead (WTD), resulting in arbitrarily large negative values: this can be redressed by a transformation of the negative values to a range with a minimum of -1 [47]. Hence the Group embarked on a programme of methodological research to embrace new methods for TTO. In addition the potential of discrete choice (DC) modelling was tested.

Time Trade-Off

In conventional TTO full health *x* is considered equal in utility terms to a given amount of time in a poor health state *t,* giving a valuation of x/t. In EQ-5D valuation studies *t* has generally been set since the early days at 10 years. Widely used in these studies, conventional TTO works well to elicit values for states better than dead. It is more problematic in eliciting values < 0. To avoid this problem more 'trading time' in full health can be provided so that, when valuing very poor states of health, respondents can trade off more time. The additional trading time can be added either before the health state being evaluated ('lead-time' TTO) or after the state being evaluated ('lag-time TTO').

Discrete choice modelling

The Group started to explore the use of DC modelling as a valuation technique for the EQ-5D in 2008 with a pilot study carried out in the Netherlands [48]. DC values broadly replicated the pattern found in TTO responses, although they were consistently slightly higher than TTO values. The main difficulty in applying DC models was that they generated values on an arbitrary scale, not on the metric of the quality (of life) component of the QALY scale. Hence DC-based values needed to be anchored on a utility scale where full health = 1 and dead = 0. It was thus decided to include a DC task in the pilot EQ-5D-5L valuation studies.

Pilot studies

The multinational study comprised a core element and experimental elements. The core pilot was undertaken in Canada, England, the Netherlands, and the US. The studies involved a DC task, a VAS valuation task, and a lead-time TTO (LT-TTO) task with 100 states (10 years lead-time duration and 5 years disease time duration).

This core study was extended to Argentina, China, Singapore and Spain. The Argentina study compared LT-TTO and lag-time TTO; the Chinese study contained an experiment with 5 years lead time duration and 5 years disease time duration; and the Singapore study contained an experiment where the time window for the disease time and the lead time were separated more strongly in the visual display. The Spanish study included a DC experiment where the paired comparisons between health states were complemented with comparisons between each health state included in the pairs and being dead, thus allowing the DC model to be anchored on a utility scale rather than on TTO data [48,49].

Follow-up studies

2 follow-up studies were conducted in the Netherlands. The first compared 6 TTO variants using an internet panel. The second study was undertaken to test the feasibility of the composite TTO approach.

Data collection

In 7 of the 8 countries data were collected in a group interview setting. In England the interviews were conducted at respondents' homes. The first follow-up study used an existing online internet panel; in the second face-to-face interviews took place in a single central location.

Valuation technology

The pilot valuation protocol was implemented in a digital setting from the outset to make standardisation of the protocol for the different experiments and languages more feasible. This used a computer-assisted personal interview (CAPI) mode of administration: the EuroQol Valuation Technology (EQ-VT), originally named the DEVT. All elements of the protocol were implemented in the EQ-VT: assigning participants to sets of states from the underlying blocked design, randomisation procedures, the iterative process in the TTO, and capturing and time-stamping the participants' responses to all tasks. Different language versions of EQ-VT were developed for which the same strict guidelines for translation of the EQ-5D instrument itself were used [29]. Lastly, the interviewer training materials were also standardised and officially translated.

Results of the programme

Results included: evidence of clustering around a small number of values, some 'short-cutting' of the valuation task, and the promising performance of the DC approach. Clustering and short-cutting also persisted in the first follow-up (internet panel) study. Hence in the second follow-up study on the feasibility of composite TTO the protocol was revised to address the various data issues. Results from the different pilot studies had clearly indicated that it was not feasible to elicit TTO values in either group interview or online settings, hence the face-to-face interview setting was adopted. In addition the lead-time approach showed that a large proportion of respondents were using the complete time scale (i.e. lead-time plus disease time) to trade off, even for states which evidently were not severe. Hence in the composite TTO study, conventional TTO was used to elicit values BTD and LT-TTO to elicit values WTD. Since the time frame for the conventional TTO was set at 10 years, it was decided to use LT-TTO with a ratio of lead time to disease time of 10 to 10 years for the elicitation of values WTD. The results from the composite TTO study showed a marked improvement in data quality, especially in respect of clustering, and fewer values WTD were recorded.

EQ-5D-5L valuation protocol

Informed by the evidence from the multinational pilot studies, the Group decided on a standardised protocol for EQ-5D-5L value set studies. The protocol centres on a systematic step-wise approach to collecting values for EQ-5D-5L health states, detailed in Table 2.

Table 2: Elements of the EQ-5D-5L valuation protocol

Start interview
1. General welcome
2. Introduction
a. Self-reported health on the EQ-5D-5L descriptive system
b. Self-reported health on the EQ-VAS
c. Background questions
3. Composite Time Trade-Off
a. Instructions and example of TTO task
b. TTO valuation of 10 EQ-5D-5L states
c. TTO debriefing/structured feedback
4. Discrete Choice
a. Instructions and example of DC task
b. DC valuation of 7 pairs of EQ-5D-5L states
c. DC debriefing/structured feedback
5. General thank you and goodbye
End interview

More detailed information on many aspects of the studies covered here can be found in [50]. Papers cover the composite TTO, the effects of lead time and visual aids in TTO valuation, lead-time versus lag-time TTO variants, dealing with 'dead' in DCE experiments, alternative specifications of TTO, and one-to-one versus a group setting for conducting computer-assisted TTO studies.

After all the detailed work put into the multi-national programme the recommended valuation methodology for endorsed EQ-5D-5L valuation studies was finally presented at the Executive Committee in December 2011.

Key outcomes of the Committee's deliberations were:
- The Executive approved the proposed protocol.
- A series of recommendations was approved, including with respect to: a warm-up task of self-rated health using the official EQ-5D-5L web version, the composite TTO, a module to obtain DCE data without duration, the composite TTO and the DCE to be administered in face-to-face interviews using the EQ-VT as a digital aid, collection of background characteristics, qualitative feedback on the valuation tasks, and sample size.
- While groups were free to experiment (e.g. VAS), TTO + DCE would always be included in valuation studies recommended by the Group for valuing EQ-5D-5L health states.

Discussion of issues raised by the multi-national programme
Oppe *et al* summarised the work accomplished in the multinational programme and raised a number of concomitant issues [46]. In summary, the protocol described above represented the culmination of an ambitious programme of research commissioned and coordinated by the Group over a 3-year period. Considerable progress was accomplished in improving the methods used to obtain TTO values, in complementing these by using additional information on preferences obtained by DC methods, and in developing explicit study designs to underpin the selection of states and tasks. For the first time, the Group had developed a standard protocol embedded in a digital aid and accompanied by interviewer training materials, which could be made available to study teams in countries wishing to develop local value sets for EQ-5D-5L. Valuation studies had already been undertaken for Canada, China, England, the Netherlands, Spain, Northern Ireland/Wales/Scotland, Uruguay, and South Korea. In preparation were sets for a number of other countries. (A further 6 were completed by December 2014).

The use of a well-described, consistent protocol across all countries would further the opportunity to compare health-state preferences between countries, and to explore the influence of cultural and other differences on health state values, an aim which echoed that of the early days.

A key methodological issue was whether computer-based technology designed to present the valuation tasks could replace the requirement for face-to-face interviews by facilitating self-completion online or in group-based settings, with the aim of reducing the costs and time involved in conducting face-to-face interviews. However, the research clearly indicated the importance of trained expert interviewers during TTO valuation tasks, and this approach was likely to remain indispensable. As the Group's work over the years had illustrated, there was no 'perfect' method for eliciting stated preferences for health states. The composite TTO had improved the means of eliciting values worse than dead, removing the need for the arbitrary re-scaling of values required by conventional TTO, but issues still remained requiring further research.

Unresolved issues also remained for DC methods, such as whether or not respondents used decision heuristics, thereby violating the DC model assumptions, and variance heterogeneity [51]. Alternative formulations of DC methods could improve the application of these methods for health state valuations, for example including duration as an attribute to be varied within the design [51,52]. Such formulations could potentially be used to provide quantitative estimates of the utility associated with different dimensions and levels of the EQ-5D-5L and to overcome the issue in DC of anchoring values at 0 and 1.

Valuation Methodology WG

As part of the strategic framework for pursuing the Group's goals a series of WGs was implemented including the VMWG, with objectives: (1) To promote methodological research to develop and test innovative valuation methods for future protocols. (2) To develop a conceptual foundation and methods for valuing bolt-ons, including the criteria for their selection and the development of the descriptive system for bolt-ons. (3) To develop a conceptual foundation and methods for valuing EQ-5D-Y.

In October 2013, subsequent to the multi-country programme detailed above that formed the basis for what can be termed *EQ-VT version 1*, the VMWG issued a request for proposals to address issues that had arisen during the programme, especially with EQ-VT, and to further develop the Group's valuation methodology.

Before this, 3 'quick fixes' were applied resulting in *EQ-VT version 1.1*. These were: (i) Advise to use small teams of dedicated interviewers, and implement extensive quality control. (ii) Reduce inconsistencies and learning effects by adding 3 practice states: one mild and one severe state, and one state that is hard to imagine. (iii) Include a prompt that checks a respondent's answer after clicking "A&B are the same" button.

Four proposals were received as a result of the VMWG's Call. These were co-ordinated in a programme which delivered a number of papers subsequently presented for discussion at the Stockholm Plenary Meeting in September 2014.

A series of decisions based on the evidence presented, and specifically related to improving EQ-VT, was made at the Executive Committee in early November 2014, then reviewed by members of the VMWG, the Value Sets WG (VSWG) and the EQ-VT support team. The final decisions at the Executive in December were, all with respect to the next version of EQ-VT: not to integrate a ranking task, with or without sorting; not to integrate the BTD/ WTD split; not to integrate non-stopping TTO; and to integrate the feedback module. This revised version was denoted *EQ-VT 2.0.*

The PRET project

Partially related to the multi-country programme a number of other significant valuation studies have been undertaken in the UK. The UK National Institute for Health and Clinical Excellence (NICE) was set up to help make better health care resource allocation decisions. It bases its recommendations on cost-effectiveness, with EQ-5D the preferred instrument for quantifying the HRQoL impact of medical interventions, using the MVH tariff. However, various developments led to the need for a re-evaluation of EQ-5D, in particular construction of the EQ-5D-5L instrument.

In order for NICE to make the most appropriate decisions, the EQ-5D population value set needed to be up-to-date, based on the latest understanding of how to value health states. The PRET project aimed to contribute towards the methodology for this revaluation. It commenced in September 2010, and was completed by March 2012. A detailed report appeared in [53].

Stage 1 consisted of an online survey with 6000 respondents. It examined key assumptions typically involved in health-state valuations through a series of binary choice exercises. Stage 2 compared the results to those of an identical survey conducted in 200 face-to-face CAPIs. Stage 3 consisted of CAPI surveys of a representative UK sample of 300, using examples of TTO, LT-TTO, and DCE with duration, each followed by extensive feedback questions. Stage 4 involved a qualitative analysis of people's thought processes during both binary choice and iterative health-state valuation exercises. It was concluded from the project's results that in order for NICE to make the most appropriate decisions, the EQ-5D tariff needed to incorporate the latest understanding of health-state preferences. Further work under the aegis of the PRET project, named PRET-AS (Additional Sample), focused on duration as an attribute of DCE: $DCE_{TTO.}$ A number of papers resulted, including a paper at the Montreal Plenary [54].

FEDEV project

The PRET work was also followed up in the FEDEV-project, funded by the EuroQol Group. Its aims: (i) to investigate the duration attribute of the DCE_{TTO} design; (ii) to examine alternative approaches to designing and modelling DCE_{TTO} data; and (iii) to examine more general questions related to the EuroQol Instrument. At the Stockholm Plenary [55], 2 papers on FEDEV were presented [56,57].

Valuation papers Plenary Meetings 2010-2014

For the sake of completeness in relation to valuation work conducted within the Group a brief summary is provided of valuation papers presented at these meetings.

Athens 2010 [58]

2 papers from the US evaluated conditional median models of EQ-5D health state preferences. Another paper considered learning effects in EQ-5D TTO valuation, and a study using US data compared hypothetical and experienced EQ-5D valuations. In addition work on the 5L instrument was reflected in a Polish study comparing 3L and 5L in a student population.

Oxford 2011 [59]

Apart from a number of papers concerning a variety of disease-specific indicators mapped onto EQ-5D, valuation papers reported on aspects of the multi-country programme, PRET, and bolt-ons.

The Netherlands 2012 [60]

With most of the valuation work focussed on the multi-country programme detailed above, two other papers concerned a multi-attribute response model, and valuing EQ-5D in Australia (both 3L and 5L).

Montreal 2013 [61]

Two complete sessions of the meeting were devoted to EQ-5D-5L valuation and DCE/TTO methodological issues. In a 'thinking out of the box' session papers dealt with the capability approach, predicting productivity based on EQ-5D, and included a debate based on an overview paper about valuing EQ-5D health states. Further papers reported on DCE with duration (PRET study), EQ-5D-5L using DCEs in Australia, and between-country heterogeneity in EQ-5D-3L scoring algorithms.

Stockholm 2014 [55]

A complete session was devoted to the ongoing EQ-VT research programme. Another session was based on DCE valuation in the PRET and FEDEV programmes. A third session comprised papers on a systematic review of exclusion criteria in national health state valuation, a value function from unsaturated valuation data sets, a Brazilian saturation valuation study of EQ-3D-3L health states, and evidence of disagreements between health scales and preferences in EQ-5D.

Evidently a considerable volume of valuation work was being undertaken, not only that associated with the various projects discussed above but also that by other members and their research programmes.

Review of EQ-5D valuation studies

Given all the valuation work being accomplished, Feng Xie *et* al have published a very useful systematic review of EQ-5D valuation studies which reviewed all existing EQ-5D valuation studies to highlight their strengths and limitations, explored heterogeneity in observed utilities using meta-regression, and proposed a methodological checklist for reporting EQ-5D valuation studies [41].
Of the 31 studies included in the review, 19 used TTO, 10 used VAS, and 2 used both TTO and VAS. Most studies included respondents from the general population selected by random or quota sampling and used face-to-face interviews or postal surveys. Studies valued between 7 and 198 total states, with 1 to 23 states valued per respondent. Different model specifications had been used for scoring. Some sample or demographic factors, including gender, education, percentage urban population, and national health care expenditure, were associated with differences in observed utilities for moderate or severe health states.

It was concluded that EQ-5D valuation studies had varied widely in their design and in the resulting scoring algorithms. Hence a Checklist for Reporting Valuation Studies of the EQ-5D (CREATE) for those conducting valuation studies was proposed: see Table 3.

Table 3: Checklist for Reporting Valuation Studies of the EQ-5D (CREATE)

Number of EQ-5D health states valued is given.
EQ-5D health states valued are described.
Method(s) of selecting the EQ-5D health states is (are) stated.
Number of EQ-5D health states valued per respondent is given.
EQ-5D health states valued per respondent are described.
Method(s) of assigning the EQ-5D health states to respondents is (are) stated.
Sample size calculation is stated.
Sampling method is stated.
Target population is described.
Rationale of choosing the target population is given.
Recruitment strategies are described.
Survey format(s) is (are) stated (e.g., face-to-face interview).
Administration of questionnaire is described (e.g., paper and pencil or computer assisted).
Preference elicitation technique(s) is (are) stated (e.g., TTO).
Method(s) of transforming raw scores is (are) stated.
Choice of transformation methods is justified.
Number of people approached is stated.
Number of respondents is stated.
Demographic details of respondents are given.
Representativeness of the respondents to the target population is described.
Definitions of inconsistent and illogical data are given.
Number of respondents included in the analysis is given.
Rationale of selecting respondents for the analysis is given.
Mean and standard error of the observed utility for each health state valued are given.
Dependent variables are stated.
Details of independent variables are given.
Details of model specification are given.
Details of model estimators are given.
Goodness-of-fit criteria are stated.
Goodness-of-fit results are given.
Robustness tests of the model are conducted (i.e., model validation).
Criteria used to choose final scoring algorithm(s) are stated.
Final scoring algorithm(s) is (are) explicitly stated.
An example of using the scoring algorithm(s) to calculate health utility is given.
Value set(s) generated by the scoring algorithm(s) is (are) given.

It is noteworthy that this is a comprehensive checklist covering both the basic design elements of any study plus detailed criteria for the conduct of modelling and scoring algorithms. This study can be linked to the role of the other WG related to valuation, the VSWG.

Value Sets Working Group

The lead author of the paper outlined in the previous section was appointed Chair of the VSWG. Evidently the Group would seek to carry forward the CREATE recommendation, as indicated by its objectives.

- To stimulate interest in producing EQ-5D (3L and 5L) value sets across the world.
- To advise and liaise with local researchers and data collection agencies in undertaking value set research using the EQ-VT valuation protocol.
- To certify value sets that follow the protocol and meet data quality/scientific standards.

There could potentially be a very wide range of countries interested in developing value sets for EQ-5D-3L and EQ-5D-5L. Creating value sets could be complex and costly, whilst coordination across countries required considerable logistical and technical skills. The Group wished to encourage the use of its valuation protocol and the EQ-VT with its associated technical, IT, and translation components. The VSWG would thus be in a position to liaise with local research teams and data collection agencies. In 2015 the activities of the VMWG and the VSWG are to be merged into the Valuation WG.

6.9 Closing remarks on valuation

The lengthy nature of this section and the detail provided are testament to the significance of valuation for the EuroQol enterprise. From the outset Group participants sought to measure *and* value health status and valuation has remained a key objective through the years. The section has traced how valuation of EQ-5D (health) states has been approached by the Group, from the initial emphasis on scaling using the VAS approach *via* the extensive efforts to operationalise the TTO method for evaluation purposes, through to the latest developments using modified TTO methods and incorporating the DCE technique.

These two approaches constitute the focus of recent and planned further valuation studies, both methodologically and empirically. Given that the Group is also concerned to develop and promote the use of EQ-5D in contexts other than economic evaluation it is likely that a re-appraisal of VAS as an approach to health status scaling will take place, along with experimentation with other techniques such as Item Response Theory (IRT). The reconstituted Valuation WG has been tasked with carrying this work forward.

7. Applications

'Applications' is a convenient term to cover the many and varied uses of EQ-5D. These cover a broad range from individual medical interventions to population health status studies. After a brief commentary on the early view in the Group with respect to applying the instrument this section then traces features and trends regarding applications through papers presented at Plenary meetings. By the time of the Biomed EQ-net programme the uses of EQ-5D had expanded considerably, hence a comprehensive list of the applications of EQ-5D was specified and this is shown. Then the disease areas initiative, in particular the disease review aspect of the activities of the CSPTF, is presented. The section concludes with data on applications based on studies registered at the Group Executive Office.

The aims and objectives of the Group as outlined at the first Plenary Meeting in Lund in 1991 were: to provide a standard instrument for describing and valuing quality of life, to elicit judgements from representative (and other) samples for cross-cultural comparison and quality weighting, and to accumulate data on these matters. There was no explicit mention at this stage of applying the EuroQol instrument in, say, specific disease areas, or in particular medical settings. Indeed priority was to be given primarily to the valuation of health states, methodological issues (reliability, validity, response rates), and comparison with other instruments. Although in the latter respect there was mention of "results in health programme evaluation", the actual "promotion of the use of EQ in health programme evaluation (in general)" was to receive no priority. With the benefit of hindsight this appears remarkable, but it is evidence that the Group was not at this stage focusing on the potential of the instrument to be a 'stand-alone' measure, especially not in health programme evaluation terms.

7.1 Application papers presented at Plenary meetings: features and trends

Two strands in application can be delineated. The first concerns studies undertaken using the EuroQol instrument in medical settings, some of which did have input from Group members, but most were accomplished by health researchers taking up the instrument. The first published application paper listed on the Group website measured quality of life in stroke [3]. The EuroQol website now lists over 4000 studies, many of them applications.

The second strand is the focus here, *viz* the application work conducted primarily by Group members and presented at Plenary meetings. This sub-section examines how the EuroQol instrument has been applied in a variety of settings, for a variety of purposes and to a range of diseases and conditions, within the context of the annual Plenary meetings. Themes include cross-country collaboration, use in

general practice and primary care, comparison of health status instruments within disease areas, patient diaries, socio-economic and health inequalities, and child applications.

Turning to specific Plenary meetings, various application features can be ascertained. The Brunel (1994) meeting included a paper which constructed *cost/QALY league tables* for a hospital in Norway using VAS values from Frome (UK) data, and TTO values from the York MVH study [62], thus exploiting opportunities for cross-country collaboration in the use of EuroQol data. Such collaboration proved to be a theme in Group work.

The Rotterdam Plenary Meeting (1997) included 2 papers which reported on a different type of application, namely the *settings* in which EQ-5D was utilised, in these cases *general practice* and *primary care* [63,64]. The first concluded that: "The results from this study showed that it is feasible to collect health status data from general practice populations, but use of this information to inform purchasing decisions requires further research." The second: "EQ-5D proves to be a practical survey tool for use in primary care with virtually no problems encountered in the course of its administration over a 5-day period (from NHS personnel or patients)."

At Hannover (1998) one paper considered the performance of EuroQol in *children* in the Netherlands, the first application with children reported at Plenary meetings [65]. 5 papers *compared EQ-5D with other instruments*. The latter papers were among the many studies over the years appearing in the literature which at the same time as applying EQ-5D to specific diseases or conditions, or in particular settings, also used other measurement instruments, whether disease-specific or other generic measures, which was one of the original expectations of the Group.

Of particular interest among the 9 applications at the Sitges (1999) meeting were another paper on the validity of EQ-5D with children, in this case asthma [66], and the use of EQ VAS in a *daily patient diary* [67].

A variety of application papers were presented at Copenhagen (2001), one of which specifically compared a condition-specific instrument (AQLQ – Asthma Quality of Life Questionnaire), a generic instrument (SF-36 – Short Form 36 dimensions), and EQ-5D as a preference-based generic instrument, concluding that it was best to include EQ-5D directly in trials, rather than infer from SF-36, for example [68]. Another paper reported on *socio-economic inequalities* in the health of the Slovenian population as measured by EQ-5D [69], the first paper at Plenary meetings to apply EQ-5D to health inequalities, again proving a precursor to further interest from the Group.

At York in 2002 application papers included: whether there were roles for EQ-5D in *delivering primary health care services* [70]; the (potential) role of EQ-5D in the *international comparison of socio-economic inequalities* in health status, based on a national survey in Hungary [71]; and a survey on tackling *health inequalities* amongst older people in the UK [72]. A further paper asked what light EQ-5D shed on *international differences in self-reported health problems* by age, sex and education level, by analysing 21 EQ-5D surveys [73]. These papers were forerunners to the interest shown in applying EQ-5D in *population surveys* which later led to the inauguration of the PHTF in 2007. Looking at the italicised features it is evident that the applications of EQ-5D had moved well beyond anything anticipated in the early days. This was confirmed by the time of the EQ-net project.

7.2 Biomed EQ-net

The EQ-net book was published in 2003 [2]. Its introduction stated that the project reflected the Group's keenness to react to the demands from users and potential users of EQ-5D for clear guidelines on how to use and apply the instrument in a variety of settings and policy contexts – listed at that time as population, clinical, and economic appraisal. With respect to applications, the project facilitated the production of guidelines for the design, analysis, and reporting of EQ-5D in clinical and economic studies, and population surveys. These were reported upon in the book, as were guidelines for alternative modes of administration *viz* proxy, observer, and telephone versions.

The uses of EQ-5D listed in the book indicate the considerable expansion in the perceived scope of the instrument that developed from the Group's initial aims and objectives:

- *Monitoring* the health status of patient groups at different moments in time, e.g. referral, admission, discharge, and follow-up of out-patients.
- *Evaluation and audit* of health care, by measuring changes in health status in individual patients and in groups of patients.
- *Assessing* the seriousness of conditions at different moments in time.
- Providing relevant information for *resource allocation* at a variety of levels.
- Assisting in providing evidence about *medical effectiveness* in processes where drugs or procedures had to be approved.
- Establishing levels of *population health status* both locally and nationally.

7.3 The disease areas initiative and the review work of the CSPTF

At the 2005 Oslo Plenary meeting the paper by Oppe *et al* was important in using stroke as an example of how to retrieve and synthesise the reported use of EQ-5D in specific diseases [74]. This was evidently relevant for the review work of the disease areas task force, to which we now turn. At this meeting the Group implemented a 'product development' strategy which had implications for the application of EQ-5D. One aspect of this strategy concerned the use of EQ-5D in different disease areas, including bolt-ons or tails to the standard EQ-5D.

The Board meeting at Oslo was also important in the applications context. The work on Cheerleader projects in previous years had indicated that it was feasible to develop substantial programmes on a few key topics, almost every Group member having been involved in application studies. However, the work undertaken was nearly always *ad hoc* and not much effort had been invested in summarising the overall experience with EQ-5D, and relating that experience systematically to the classifications of severity of disease used in any specific area. Furthermore, this exercise had rarely been extended to develop a variant of the EQ-5D which would be as efficient in generating information as the EQ-5D, use the basic questions of the EQ-5D, and provide vital information on aspects of disease areas which were not covered by the standard questions of the EQ-5D. The purposes of the disease areas programme would thus include: (i) To produce summaries of the results of EQ-5D in different disease areas, often referred to as 'reviews' in subsequent work. (ii) To ascertain whether there was potential to develop a disease-specific short tail, to test it and field in a pilot study. (iii) To retrieve data for meta-analyses.

With respect to the disease area programme (eventually re-named the CSPTF), the rest of this sub-section will focus primarily on the review aspect of the task force's activities. Reviews funded by the Group were undertaken on cancer, cardiovascular diseases, cerebrovascular disease/stroke, cost-utility analysis applications, diabetes, injury and trauma, multi-morbid conditions, osteoporosis, asthma, and chronic obstructive pulmonary disease (COPD). Reviews not funded by the Group but involving EQ members included HIV/AIDS, schizophrenia, and visual disorders. In addition a number of reviews were published that did not involve EQ members: liver disease, cardiac arrest, glaucoma, generalised anxiety disorder, and HRQoL and resource allocation in children.

At the 2006 Barcelona Plenary 2 papers were products of the disease areas funding programme: (i) The use of EQ-5D in cost-utility analysis, covering 45 studies [75]. (ii) The use of EQ-5D in cancer, with 31 articles reviewed, subsequently published [76,77]. The 2007 Netherlands Plenary included a paper reviewing the use of EQ-5D in COPD and asthma, also subsequently published

[78,79]. It was noted at this meeting that the task force had developed a template for reporting EQ-5D in different disease areas, and that its scope had been widened to include a repository of patient-level EQ-5D data and the parallel field testing of the 3L and 5L versions in patient sub-groups.

The CSPTF issued a progress report in February 2009. After noting that the task force had continued to concentrate on summarising the literature on specific applications of EQ-5D, it reiterated that future work would focus on establishing a repository of patient-level EQ-5D data, and the parallel testing of EQ-5D-3L and EQ-5D-5L versions in patient sub-groups. Later that year the Paris Plenary included one of the review papers resulting from the CSPTF's projects, on type 2 diabetes mellitus, covering 59 studies for the period [80].

The Athens Plenary (2010) included a review paper of 31 studies over the period 1990-2009 which asked whether EQ-5D was inferior in content to disease-specific utility measures [81]. At the Oxford Plenary (2011) there were 2 review posters, one on the performance of EQ-5D in 4 disease areas, the other on EQ-5D and SF-36 in mental health [59].

The final CSPTF Report, in September 2011, provided a listing of post-2009 publications on reviews. These covered diabetes [82], cardiovascular disease [83,84,85], and injury [86].

7.4 Later Plenary meeting applications

Evidence that the newly constructed EQ-5D-5L was beginning to be applied came at the Barcelona Plenary (2006), which included 2 papers on the 5L version with specific patient groups, namely chronic lymphocytic leukaemia and cancer [87,88]. This was followed at the Athens Plenary (2010), where 2 papers reflected trends in EQ-5D research with some relevance in applying the instrument. First, a paper from Italy compared the new 5L version with the 3L instrument in patients with liver disease [89]. Second was a pilot bolt-on study testing a version of EQ-5D-5L specifically for use with people with psoriasis, EQ-5D-Psoriasis [24].

At the Oxford Plenary (2011) a poster considered patient-reported outcome measures (PROMs) in assessing quality of care in patients with initial knee and hip replacement in German hospitals [90], and a session on comparison of 3L and 5L included a paper comparing EQ-5D-5L to EQ-5D-3L in 8 patient groups [91]. The Netherlands Plenary (2012) included a paper on PROMs in Slovenia involving 4 surgical procedures [92], and another on PROMs in knee replacement in the UK [93]. PROMs studies continued at the Montreal Plenary in 2 papers, one concerning hip replacement

surgery in the UK [94], and the other from the USA using EQ-5D-5L and the Patient Reported Outcomes Measurement Information System 43-item short form (PROMIS-43) in COPD [95].

7.5 Registered studies

In 2009 the Group requested users and potential users who were applying to use the Group's instruments to register their studies. By the time of the completion of the book on the Group's history some 6800 studies had been registered [1]. Listed in the book were the extensive range of clinical areas (over 80) and the broad range of programmes and settings (40) in which EQ-5D had been employed. For studies in the top 25 clinical areas, apportioned by type of study the figures were: randomised control trials 29%, observational studies 27%, surveys 27%, and other studies 17%. Programmes and settings included, *inter alia*, surgical procedures, general practice and primary care, hospital waiting lists, physiotherapy, and rehabilitation. These studies were undertaken in a wide range of countries. It was noted in the book that "with the large number of language versions now in use across the globe there is no question that applications of EQ-5D will continue apace." This was confirmed by the rise in the number of registrations to over 12000 in 2014.

The application of EQ-5D in the population context has been detailed in book form [96]. A follow-up to an earlier work [97], the book presents population norms for 24 countries and some of their regions, as well as cross-country analysis and socio-demographic indicators.

7.6 Conclusion on applications

It is evident that EQ-5D has been utilised across a wide range of diseases and conditions, in a variety of settings and programmes, and in a large number of countries. It has also been used in conjunction with, and compared to, a large number of other health and health-related instruments, both generic or multi-attribute and disease- or condition-specific. The reference section of the EuroQol website continues to reflect the increasing volume of publications using EQ-5D. Finally it can be noted that application activities have been part of the remit of the Large Scale Applications WG and the Data Archive Working Group, whose roles are covered in the next section.

8. Product development and research strategy

8.1 Introduction

The well-financed and highly structured Biomed EQ-net approach which made up the bulk of the Group's work around the turn of the millennium provided the impetus to apply a more programmatic approach to the Group's scientific activities. Another factor helping to underpin this approach was the steady increase in finance accruing to the Group as the usage of EQ-5D increased, not least as it was translated into more languages.

This section traces the development of the strategy and the opportunity is also taken to explore the activities of the Task Forces not hitherto covered in detail in previous sections, namely the DTF, the YTF, the PHTF, and finally the system of WGs which replaced the task forces in 2012/13 is outlined. This system was itself being reorganised as a consequence of the latest *Statement of Research Priorities* issued by the Board and Executive Committee in January 2015.

The Board decided in 2001 that the Foundation should fund scientific research studies, in part to build upon the EQ-net achievements. Key research topics were identified and members offered to take the lead on each of these topics ('cheerleaders').

The term 'product development' was first used formally within the Group in 2004. At the Executive's September meeting that year it was noted that EQ-5D had been very successful during the period 1993-2003, with its diffusion worldwide exceeding expectations. In the previous 10 years, product development had primarily focused on the production of multiple language versions of EQ-5D, by then more than 70 in total. The Group could now decide simply to consolidate its achievements and remain an organisation that supported the use of EQ-5D by focusing primarily on academic and methodological issues. This would be a waste of the joint knowledge and varied expertise of members, which instead could be channelled into the further development of EQ-5D. Potential products were in the areas: increasing the levels of the EQ-5D descriptive system, different modes of administration, EQ-5D in specific population groups (e.g. children), and digital representations of EQ-5D.

In 2005 the Executive Committee decided that key areas where research should be undertaken were: (i) increased levels for EQ-5D (including the labels); (ii) disease areas (use of EQ-5D in different disease areas, bolt-ons to standard EQ-5D), and (iii) valuation methodologies. These would be the remit of the task forces which were subsequently established.

This list indicates the comprehensiveness of the approach *viz*: Valuation (VTF), Digital (DTF), Condition-Specific (CSPTF), Increased Level (Labelling), Youth (YTF), and Population Health (PHTF). The VTF, CSPTF and Increased Level task forces were heavily involved in issues - the descriptive system, valuation, application - which have been dealt with in some detail earlier in this paper, so need no separate treatment. More 'stand-alone' were the DTF (although valuation aspects meant close interaction with the VTF), the YTF (which also liaised with the VTF), and the PHTF; hence their activities are outlined separately below.

Concurrently with the task force programme the Group also wished to encourage members to undertake innovative research so in 2007 the Executive decided to set up a mechanism to deal with calls for proposals, identifying two types of projects - those that fitted within the scope of the task force programmes and those within a wider innovative programme.

The Executive's meeting in November 2009 was notable for the formation of a small English language reference group to advise on version management. This VMG replaced the Translations Committee and was strongly involved in further product development, as reported in section 5.

In 2011 the Executive reviewed scientific priorities and budget allocations to principal research areas, and endorsed a recommendation to the Board to make financial reservations for: a programme of 15 valuation studies of the 5L, a programme to develop bolt-ons, and research in the area of the Youth version. Valuation and bolt-ons have been treated previously, and the Youth programme is considered later in this section, along with the PHTF's activities. First, the digital strategy pursued by the Group is considered.

8.2 Digital strategy

The product development strategy was significantly influenced by the burgeoning innovations of the digital age. In response the Group launched a Digital Squad. Its inauguration took place after a number of years' engagement by the Group on electronic developments.

The phrase "computer-interrogation" was actually used at the 1st meeting in 1987! There was a section on computer-assisted data collection in the EQ-net book which clearly showed that the group was aware of the potential for computer use and the challenges posed in delivering electronic versions of EQ-5D consistent with the pencil-and-paper version [98]. Computer-assisted modes of administration were likely to play an increasing role in survey data collection, with the potential for reducing the resources spent in time, labour and checking for error in the manual coding of paper-and-pencil questionnaires, data-entry, and data-analysis. In theory, the use of computer-assisted methods should

introduce little bias in terms of framing effects. Depending on the design, computer-assisted tools were just screens which should display the same stimuli (i.e. sections of the EQ-5D) as those seen on paper, and to be consistent, the computer screen should display the 20 cm vertical EQ VAS. Of interest, given the later deliberations of the DTF, was that it was recommended that the format of computer-assisted data collection versions (i.e. the wording and layout) should be exactly the same as the corresponding recommended pencil-and-paper version. Standardised instructions should be written and a standardised screen used.

In the same year (2003) as the EQ-net book's publication the Board discussed a potential EuroQol Electronic Programme (EEP) which was intended to provide a framework for the development and funding of EQ-5D software proposals. Although this EEP was not formally implemented the Board and Business Management continued to work on digital issues. In 2006 a draft contract was signed between the Group and a company specialising in IVR (interactive voice response) scripts that would enable the production of EQ-5D IVR scripts.

Next the Digital Squad (later, DTF) was inaugurated in May 2008 to provide a forum to consider the challenges, issues, and benefits of digitisation. Much of the DTF's work related to how to deal electronically with the descriptive system and EQ-VAS. Once in place the digital products were employed in the service of the valuation work, in particular the EQ-VT programme. The latter activities were described in section 6.8.

At the 2011 Oxford Plenary it was agreed that the DTF would enter a phase of suspended animation, and that the role for monitoring electronic products would henceforth be undertaken by the VMG with its wide remit to manage all versions of EQ-5D.

In conclusion it can be observed that the originally proposed EEP approach eventually developed towards advanced electronic valuation software. It can be claimed that the EuroQol Group is now leading in this electronic valuation technology, not for DCE which can be done by marketing panels, but for TTO which is more demanding. It is evident that the expansion of digital technology has greatly increased the scope for collecting information on HRQoL from respondents employing the various versions of EQ-5D. The 'paper and pencil' approach has not been abandoned. To do so would run the risk of side-lining the considerable repository of EQ-5D data. Finally, it should be reiterated that the work of the DTF was built upon in the EQ-VT framework currently utilised for valuation purposes.

8.3 Youth Task Force

The Child-Friendly TF was inaugurated in 2006, then re-named the Youth TF in November 2008 following the Executive Committee's selection of EQ-5D-Y as the designated name for the child 'product'. This task force contributed substantially to the activities of the Group as a significant number of members was involved.

It is first worth considering the background to the establishment of the task force. A number of papers were presented on the use of EQ-5D with children and students at Plenary meetings between 1998 and 2001 [65,66,99]. The first attempt to provide what came to be called a 'child-friendly' version of EQ-5D was reported in 2002 in a paper which was to have considerable influence on subsequent work in this area: Hennessy and Kind tried out what they termed a 'modified wording' version of EQ-5D in a sample of secondary school children (aged 11-16) in England [36]. This was followed by a series of papers and posters from a range of countries at the Bled (2003), Chicago (2004), and Oslo (2005) Plenary meetings.

Inaugural Meeting of the Child-Friendly Task Force January 2006
A number of conclusions with regard to the work of the task force were drawn: (i) The construction of a standardised version that could be modified in other countries. (ii) There was agreement that a child-friendly EQ-5D version should be developed, but not yet about age ranges. (iii) It was important to investigate the validity of such a child-friendly version. After reviewing the UK version the outcome was a proposed international child-friendly version.

Deliberations at subsequent meetings can largely be framed in a series of themes: the youth instrument itself and its administration, dimensions, proxy versions, age ranges, the VAS, expansion to 5L, valuations, and publications. These are each briefly treated.

The youth instrument and its administration
The Task Force meeting in September 2007 decided that priorities for studies should include: (i) Patients with health problems, bigger samples/ proportion of sick, cultural differences, and longitudinal or interventional studies. (ii) Pilot studies with different wording: timing (i.e. 'your health today'), problem description (i.e. asking whether problems were health-related), (iii) Different methods of data collection (pencil, internet, phone etc).

In 2008 the designation 'EQ-5D-Y' was selected for the youth instrument and the task force renamed the Youth Task Force. Its meeting in May 2009 gave final approval to the five existing language

versions. The Executive Committee Meeting in the same month proved a landmark for the YTF and the Group when EQ-5D-Y was approved as an official EQ-5D product. The YTF was the forum for deliberation on the youth work until it was superseded in 2013 by the EQ in Children WG.

Dimensions

The dimensions of EQ-5D-Y were the subject of task force deliberations in two major respects: the content and labelling of the dimensions themselves, and the possibility of adding more dimensions to the instrument. The outcome of the YTF's work was the EQ-5D-5Y with the following dimensions: *mobility (walking about); looking after myself; doing usual activities*; *having pain or discomfort*; *feeling worried, sad or unhappy*. Thus the 5 dimensions from the adult EQ-5D remained in the Youth version EQ-5D-Y, but the dimensions were renamed in child-friendly wording to make them easier to understand for children and adolescents.

As with the main EQ-5D, the dimension issue in the youth work has mainly focused on the potential for bolt-ons. Thus at the Oxford (2011) Plenary a poster was presented on developing a cognitive dimension as a bolt-on for the EQ-5D-Y in Germany [100]. The responsibility for the development of bolt-ons was then transferred to the EQ-5D in Children WG.

Proxy versions

Proxy versions were first mentioned in the task force context in 2006, agreement on the final version being reached in November 2010. A series of studies was undertaken involving alternative proxy versions.

The YTF reported at the Oxford (2011) Plenary Meeting, where the instrument was shown. The current position with respect to proxies is given in the EQ-5D-Y *User Guide*. EQ-5D-Y has two proxy (source) versions:

Proxy 1: The proxy rates how he/she rates the health of the child. It can be used for children from 4-7 years and for children of 8 and over who are not able to fill in the EQ-5D-Y themselves.

Proxy 2: The proxy rates how he/she thinks the child would rate his/her own state if he/she were asked directly and could communicate it.

The EQ in Children WG continued to focus on the validation of EQ-5D-Y in younger age groups using the proxy version, an example being work on the validity and reliability of the Spanish EQ-5D-Y proxy version presented as a poster at the Montreal Plenary [101].

<u>Age ranges</u>

The age range deemed appropriate for child-friendly versions of EQ-5D was an issue from early in the task force's deliberations. In 2007 it was reported that studies had been undertaken in Germany, South Africa, and Spain. The evidence suggested an EQ-5D (Child) version was generally recommended for children aged 8-16 years. In detail:

- Age 0-7: no EQ-5D.
- Age 8-11: EQ-5D (Child).
- Age 12-15: overlapping area. Generally EQ-5D (Child) was recommended. However, depending on study design the usage of the EQ-5D adult version might be possible.
- Age 16 and older: adult version. A possible exception applies for studies involving only children up to 18: EQ-5D-Y would be recommended to avoid any potential discontinuity between use of the youth and adult versions, as they are different instruments.

This basic framework subsequently remained the same, except 'EQ-5D-Y' replaced EQ (Child). It is shown in the *EQ-5D-Y User Guide*.

The YTF Report at the Oxford (2011) Plenary noted that among the challenges for EQ-5D-Y valuation was evidently the different age groups involved. The YTF was exploring valuation methods for the EQ-5D-Y version in interaction with the VTF.

One of the objectives of the EQ-5D in Children WG was to build upon the work of the YTF to develop EQ-5D instruments suitable for use in children of various age ranges, including the validation of the EQ-5D-Y in younger age groups (using the proxy version).

<u>VAS</u>

The design and wording of the VAS was raised in 2006, and after some years of variations in practice across languages, in more detail in 2008. In respect of VAS-Y the format of the different versions should look as similar as possible and only deviate if there were strong reasons for recommending changes. At the YTF Meeting in February 2009 a number of decisions were made with respect to the VAS: the scale should be retained, the endpoints of the scale should be 'the best health you can imagine' / 'the worst health you can imagine', and marking with an X should be retained. This text would be translated into Swedish, Spanish, Italian, and German.

Expansion to 5L

A 5L version was first mooted in 2006. Since work was proceeding on the adult 5L version by the time of the September 2007 YTF Meeting it was recommended to wait for this adult version. The task of developing the EQ-5D-5L-Y was then taken up by the EQ-5D in Children WG.

Valuations

The important issue of children's value sets was first treated in December 2006. A number of questions were raised which re-surfaced in the YTF in 2009. The key question concerned whether a separate Youth value set was needed and, if so, the methodology to be used given that the VTF was exploring different methods. Points raised Included: (i) There was no available information about hypothetical health states among children. (ii) Could children cope cognitively? (iii) For clinical studies 'own health' ratings might be used, but for economic studies adult general population samples could be asked about values for children. (iv) Should children be treated differently?

At the VTF Meeting in September 2009 it was decided that a liaison group would be formed to interact with the YTF on children's values. From 2010 onwards the YTF continued to deliberate on valuation, and in a report of January 2012 from the YTF Chairman it was recorded that the YTF should keep the main responsibility for the pilot valuation studies. One pilot valuation study was completed: the main result was (based on VAS measurement) that adults value health differently comparing adults (themselves) and a hypothetical child. The other, using TTO/DCE measures, is mentioned below. Further valuation work was taken up by the EQ-5D in Children WG and this is also outlined below.

Summary of child-friendly work under the aegis of the YTF

The child-friendly work resulted in a number of papers and publications. Most of these were the result of co-operative endeavours under the aegis of the YTF. After a prolonged process dating back to 2007 two papers were published which are worth summarising to indicate the scope of the YTF's work. The purpose of the first paper (Willie *et al*, 2010) was to develop a self-report version of EQ-5D for younger respondents, EQ-5D-Y, to test its comprehensibility for children and adolescents and to compare results obtained using the standard adult EQ-5D and the EQ-5D-Y [102]. The YTF revised the content and wording of EQ-5D to ensure relevance and clarity for young respondents. Children's and adolescents' understanding of the EQ-5D-Y was tested in cognitive interviews in German, Italian, Spanish and Swedish. Differences between the EQ-5D and the EQ-5D-Y regarding frequencies of reported problems were investigated in Germany, Spain and South Africa. The content of the EQ-5D dimensions proved to be appropriate for the measurement of HRQoL in young respondents. The wording of the EQ-5D questionnaire had to be adapted, leading to small changes in the meaning of some items and answer options. This adapted EQ-5D-Y was satisfactorily understood by children and

adolescents in different countries, was better accepted, proved more feasible than EQ-5D and was thus a useful tool to measure HRQoL in young people in an age-appropriate manner.

The second publication examined the feasibility, reliability, and validity of EQ-5D-Y [103]. It was administered in population samples of children and adolescents in Germany, Italy, South Africa, Spain and Sweden. Results included: (i) Between 91% and 100% of respondents provided valid scores. (ii) Percentages of agreement in test-retest reliability ranged between 69.8% and 99.7% in the EQ-5D-Y dimensions. (iii) Correlation coefficients with other measures of self-rated health indicated convergent validity. (iv) Differences between groups classified according to presence of chronic conditions, self-rated overall health and psychological problems provided preliminary evidence of validity. Overall, the study's results provided preliminary evidence of the instrument's feasibility, reliability and validity.

EQ-5D in Children Working Group

Under the Working Groups strategy the YTF was superseded by the EQ-5D in Children WG in order to build upon the work of the YTF in developing EQ-5D instruments suitable for use in children of various age ranges, including the validation of the EQ-5D-Y in younger age groups (using the proxy version).

Its objectives were to develop and test a 5L version of the EQ-5D-Y, to promote research exploring the validity of the EQ-5D-Y as a measure of health status in children, to work with the VMWG in developing a programme for the valuation of EQ-5D-Y states, to develop and to test possible bolt-ons and bolt-offs (with regard to the self-care dimension) for the EQ-5D-Y in co-operation with the VMWG, and to develop a user guide for the 3L version of the EQ-5D-Y.

Taking the latter first, the *User Guide* was published in August 2014, and is available on the EuroQol website, which lists 34 language versions of the EQ-5D-Y self-complete version.

The research programme of the WG aimed to identify appropriate level labels for a 4L or a 5L version of EQ-5D-Y, and to test the comprehensibility and feasibility of alternative 4L or 5L versions to obtain the new instrument. The research design initially involved the development of a pool of possible labels, and testing and rating of identified labels by conducting interviews using response scaling tasks. Analysing the results from the response scaling tasks would be likely to identify two alternative five-level versions (as happened in the adult 5L work), or a 4L and a 5L version. In phase 2 these versions of EQ-5D-5L-Y were to be tested in the relevant age group (8-15 years) by cognitive interviews with children and adolescents.

In March 2015 the label extension study is running, in England, Sweden, Spain, and Germany, with results to be presented at the Plenary Meeting in Krakow. With respect to the valuation of EQ-5D-Y, the pilot valuation study noted above is being followed by another pilot study but involving TTO/DCE measurement. Study teams from England, The Netherlands, Spain, and Germany are involved.

In 2015 the WG has been re-named the EQ-5D in Younger Populations WG.

Conclusion

This section has dealt with the child-friendly aspects of the Group's activities in a little detail. It is evident that this area has excited the interest of a significant number of the Group's members, which should not be surprising given the importance of the health status of children in all societies. The work of the YTF and now the WGs has involved some very interesting issues for the EuroQol Group in respect of, *inter alia,* the adaptation of EuroQol dimensions, labelling, the development of proxy versions, and valuations: this can be expected to remain the case.

8.4 Population Health Task Force

This task force was set up in 2007 and held its first meeting in November of that year. Its aim was to promote the use of EQ-5D in population studies by encouraging awareness and informed use of the EQ-5D in population health applications. The number of Group members present (13) at the 1st PHTF Meeting in November 2007 indicated the level of interest. Population health work using EQ-5D either being undertaken, or proposed, was reported for Sweden, Spain, Italy, Norway, US, Canada, and China.

Potential activities for the task force were listed, including: develop new external associations (epidemiology, health policy, medical demography, public health medicine), re-launch EQ-5D as a summary measure of population health, seek national or external funding for major new population health surveys, sponsor new population health surveys in countries of interest.

It was agreed that task force members would provide an overview of work in the area of population health using EQ-5D, according to the following explicit criteria: study objectives/ design; methods – sampling frame, survey methodology; data – results; study limitations; format; mode of administration; contextual variables (gender/education/income/housing); publication source; sponsorship/funding.

At its September 2009 meeting PHTF members reported on a range of population-related studies from Spain, Greece, Italy, Sweden, China and Canada. Another study covered 4 countries: Chile, UK, Turkey, US.

The PHTF next reported on its activities in February 2010: (i) An overview of the use of the EQ-5D in health surveys. (ii) Proposals for revision of the work from Sweden, UK, Italy, Canada, Finland, Spain, and North America were funded by the Executive Committee. (iii) A template/ data abstraction form was agreed to standardise these revisions. (iv) Revisions from 5 countries (14 health surveys) had been completed by the end of 2008, and summary tables of the main data were produced in 2009. Under the aegis of the PHTF, the Population Norms Project produced an updated edition of the population norms book [97], published in 2014 [96].

The task force continued to deliberate its research agenda and the challenges it faced, but when the task forces were discontinued most of this agenda was transferred to the Large Scale Health Applications WG.

8.5 EuroQol Working Groups

In 2012/13 the Executive decided to introduce a new structure comprising 6 WGs, with clearly defined and specific objectives relating to the scientific agenda of the Group. They built either directly on previous task force work and/or covered entirely new activities or areas of enquiry. They were: Valuation Methodology (building on VTF, DTF, CSPTF), Value Sets (VTF, DTF, CSPTF), EQ-5D in Children (YTF), Data Archive (CSPTF), EQ-5D in large scale health applications (PHTF), EQ-5D beyond health care. These groups were designed to be the principal agencies to channel ideas concerning research priorities and requests for proposals.

Three key considerations were involved: (i) An effective way was required to generate more high quality research proposals focussed on the scientific agenda. (ii) Ensuring all EQ members were given the opportunity to be involved in research activities coordinated by WGs by establishing Special Interest Groups (SIGs) comprising Group members with demonstrable interests in the relevant areas. (iii) It was essential to retain and promote new ideas and fresh thinking, and to support innovative research not identified as a priority. Hence there would be renewed efforts to invite proposals on innovative topics via regular requests to the wider membership.

The VMWG and VSWG have been considered in section 6, and the EQ-5D in Children WG following the section on the work of the YTF. The roles of the Large Scale Health Applications WG, the EQ-5D beyond health care WG, and the Data Archive WG are briefly considered here.

Large Scale Health Applications WG (LSHA WG)

The objectives of this WG were to stimulate interest: in routine use of EQ-5D in large-scale health applications, in methodological and applied research relating to the use of EQ-5D in measuring provider/health care system performance, in methodological and applied research relating to the use of EQ-5D in assessing populations; and to explore the development and use of new and existing EQ-5D products for large-scale health applications.

In its call for proposals in April 2014 the LSHA WG pointed to the wealth of population survey data that incorporated EQ-5D alongside other measures of potential interest, for example life satisfaction, happiness, subjective well-being and an wide array of personal/demographics covering marital status, education, economic status, income and accommodation. It also noted that EQ-5D could be found embedded within large-scale information systems where it formed part of clinical/administrative records. The volume of these datasets provided the Group with new opportunities to explore the relationship between EQ-5D and a range of social-economic and health variables that were of relevance to policy makers and to those who plan, commission and deliver health services. Hence the LSHA WG welcomed new proposals showing how existing data could be exploited to create a new generation of indicators with potential value to decision-makers at all levels within health care systems and possibly beyond. This WG is to continue within the new framework set up in 2015.

EQ-5D beyond health care WG

The objectives of this WG were to explore the use of EQ-5D as a measure of outcomes in interventions outside the health care sector e.g. social care, housing, residential care, impacts on carers, and to ensure the Group developed the knowledge and evidence required to respond to challenges arising from other measurement approaches e.g. capabilities, subjective wellbeing. This WG is to be discontinued in 2015, with future initiatives in these areas to be considered as innovative research.

Data Archive WG (DAWG)

From early in the Group's existence it was realised that, given the volume of empirical work being undertaken, there was scope for data pooling in archival form which could enhance the potential for EQ-5D research. Data archiving material generated in the early years was transferred to the Business Management in 1994.

A few years later the main thrust of the EQ-net project with respect to data bases was the construction and analysis of harmonised VAS and TTO data bases [104]. Resources were made available to the Business Office in 2001 to continue to analyse the data emanating from these harmonised European

datasets. Over subsequent years the databases were maintained, especially the valuation sets, and the latter were used in the value set book published by the Group [39].

In 2008 the Group started exploring the possibilities of storing large datasets. Negotiations with academic users engaged in large studies (N≥ 5000) proved successful and the first datasets were added to the data archive. The total number of observations for signed contracts by early 2012 was over 500000.

The mission of the DAWG was to provide a structure and process for managing research using existing data sources that included the EQ-5D, where there was an executed contract containing a data sharing agreement between the study investigators and the EuroQol Business Office. Its primary aims were: (i) To provide information/summarise the data from studies registered with the EQ office where a sponsor had agreed to potentially share the data. (ii) To facilitate interactions between EQ members and trial sponsors with regard to data analysis and resultant publications.

The Key Databases proposed by the DAWG in early 2013 were DB1 Study Meta-Data, DB2 Patient data, DB3 Core Measures, and DB-S Supplemental Databases. By April 2014 the DAWG had archived datasets from 27 different studies. These were categorised as: (1) *Large User Databases:* data from large studies (n ≥ 5000) conducted by academic and non-academic users, i.e. non-pharmaceutical companies/medical device manufacturers or any other for-profit making stakeholder; (2) *EQ-5D Funded Studies;* (3) *Non-EQ-5D Funded Studies;* and (4) *Publically Available Datasets*. In 2015 the tasks of this WG are to be integrated with Business Office activities.

8.6 Revised Working Groups Framework

A number of mentions have hitherto been made of changes to WGs. As a result of a new research strategy to be implemented in 2015 a revised framework of 6 WGs is being established which is linked to the Group's Strategic Research Priorities: see Table 4.

Table 4: Revised Working Groups

2012-14	Proposed 2015-2017	Comment
	Descriptive Systems WG	This new WG will focus on exploring the conceptual basis for generic preference-based HRQL measures.
Value Sets WG	Merged	The activities of this WG will be incorporated into the Valuation WG.
Valuation Methodology WG	Valuation WG	Continues to focus on valuation methods. However, the activities have been extended to include international initiatives related to prospective valuation studies and international initiatives using data collected from valuation studies.
EuroQol Data Archive WG	Integrated with business office activities	The activities will be integrated with business office under the supervision of a senior scientist, and the existing DAWG members will be asked to serve as an advisory board to provide guidance to the office.
EQ-5D in Children	EQ-5D in Younger Populations WG	Continues.
EQ-5D in large scale health applications	Large Scale Applications WG	Continues; modify aims to support clinical and population based initiatives.
EQ-5D beyond health care	Discontinued	Future initiatives will be considered as innovative research.
	Interface Development WG	New WG; will focus on new applications of existing instruments, ways to collect data, social media (the D in R&D; more development than research); potential for commercialization and commodification will be considered.
	Education and Outreach WG	As part of the broader mission of the EuroQol group, this WG leads initiatives to educate members, and the broader scientific community/ policy maker to promote better understanding of the uses of EQ-5D and its underlying science.

8.7 Research themes and priorities

Also as a consequence of the latest strategic developments a series of research themes and priorities has been determined.

(i) Explore the conceptual basis for generic preference-based HRQL measures.

(ii) Investigate new approaches to valuing health (not necessarily associated with the conventional QALY paradigm).

(iii) Examine large scale health systems applications for EQ products.

(iv) Support valuation studies for the EQ-5D-5L. (a) Support the development and dissemination of EQ-5D-5L value sets in key regions. (b) Explore valuation research in specific patient groups.

(v) EQ-5D-Y: refinement of descriptive systems, valuation studies.

In addition, innovative research outside the scope of the above areas continues to be encouraged, including, e.g., bolt-on studies.

8.8 Concluding remarks on research strategy

Since the Group became more explicit in its strategic approach following the EQ-net project it has progressed through a series of developments which have led to the latest structure of Working Groups and the articulation of research themes and priorities just outlined. The Group's strategic approach can be expected to adapt to the changing circumstances of HRQoL measurement and valuation and the demands of the health and related sectors worldwide.

9. Summing Up

28 years ago a small group of people from a variety of disciplinary backgrounds met to discuss the feasibility of developing an instrument for the measurement and, crucially, valuation of health status. Uppermost in the initial discussions were the beliefs that the instrument should be kept as simple of possible, enabling its self-completion and acceptability for use in postal surveys, and that it should be a standardised tool which would be amenable to the pooling of data from studies conducted by group members, in the early days based in a small number of institutions. Considerable time was taken on developing the instrument in such a way that there was a 'common core' of data that could be so pooled. From these relatively humble beginnings the EuroQol instrument has now reached global status having been translated into many languages and been utilised in a great variety of applications by a wide range of organisations.The paper has endeavoured to provide a concise history of the

Group's activities up to the present day whilst preserving some depth on each major aspect of the Group's activities.

A key influence on the Group's activities was the (largely unexpected) and quite rapid take up and use of pages 2 and 3 of the EuroQol instrument, to the exclusion of the valuation elements of the questionnaire, by medical and health sector personnel from the early 1990s.These pages were originally intended as a 'warm-up' exercise to enable respondents to understand the nature of the valuation task they were being asked to undertake. Another original expectation was that the instrument would be used alongside other health status instruments and disease/condition-specific measures in evaluating health programmes. The latter certainly proved to be the case, but it was as a 'stand-alone' instrument, termed 'EQ-5D' from 1995, that its dissemination was most rapid. Reasons for this expansion in the use of EQ-5D included its relative simplicity - at just 2 pages it proved user-friendly, its reliability, and its ability to generate numbers to characterise health states.

The success of the EuroQol Group has also been closely related to developments in health policy – first the implementation of health technology assessment (HTA) based on cost per QALY analysis, and second the influence of HTA continued to expand internationally. Other uses for measures of health, such as in the routine outcome measurement of patient reported health to monitor the quality of health care providers, and the use of patient's views about their health in shared decision making, also emerged.

Another aspect of the expansion in the utilisation of EQ-5D, geographically and otherwise, was the 'open door' policy pursued by the Group from the outset. This included publishing the proceedings of Plenary meetings, which are all available in both book form and on the EuroQol website, and extended to an open access policy for Group publications.

After the substantial resources made available for the Biomed EQ-net project, a pricing policy with respect to licensing the use of EQ-5D was pursued which charged appropriately for commercial utilisation, but allowed free usage for clinical and academic research not funded by a pharmaceutical company. This policy supported what has always been a small-scale central (business) operation, but with the financial resources that have sustained and latterly expanded the activities of the Group with members dispersed across a wide range of institutions. The pricing policy was modified to a more nuanced approach involving 5 main user categories in 2008, but evidently remains acceptable to users since licensing of EQ-5D products continues to be sustained.

Another key feature was that whilst increasing the number of levels was debated off-and-on at Plenary meetings, and occasionally additional dimensions were under consideration, EQ-5D remained largely

unchanged (apart from some minor alterations in design and wording) for over 15 years. This was partly in reaction to the heavy intellectual and resource investment into developing the instrument in the first place, and partly for good scientific reasons connected with, e.g., the richness of the data accumulated over time and hence the potential problems in comparing and perhaps integrating this data following any substantive change. And, in any case, EQ-5D was proving to be very successful! Eventually the weight of evidence, along with expanding activity, e.g., in measuring the health status of children and young people, led to additional instruments in the Group's 'portfolio', namely EQ-5D-5L and EQ-5D-Y. This was consistent with a tenet in place from the early days of the Group's efforts: any decisions should only be taken following the empirical testing of any proposed change or innovation.

This paper has not provided a critical review of the EuroQol Group's activities. Criticisms raised over the years have included specific points with respect to EQ-5D which have emerged in empirical work such as lack of sensitivity and ceiling and floor effects. Attention paid to such criticism did lead, however, to changes, e.g. the construction of EQ-5D-5L. The relatively simple nature of the instrument has led some analysts to conclude that it is too 'simplistic' in the sense that it does not capture 'enough' aspects of health status. This point has been (partially) addressed in the development of bolt-ons, but it should be stressed that the original aims of the Group led to the development of an instrument which purposely minimised the number of dimensions and of levels and thus accepted the trade-off between simplicity and comprehensiveness.

Other criticism has been of a more wide-ranging kind to do with the wisdom, or otherwise, of characterising health in quantitative terms. Indeed in some countries, e.g., the QALY has not been approved for use in health sector appraisal.

Finally, the valuation of health remains complex and, at times, controversial. The Group continues to undertake substantive work in addressing this complexity, and has been much encouraged by the wide geographical spread of this work, and of all the wider activities of the Group, to an extent simply not foreseen in the early days by those who named it *Euro*Qol.

References

1. Brooks R. The EuroQol Group after 25 years. Dordrecht: Springer, 2013

2. Brooks R, Rabin R, de Charro F. The Measurement and Valuation of Health Status using EQ-5D: A European Perspective. Dorderecht, Kluwer. 2003.

3. de Haan R, Aaronson N, Limburg M, Hewer R, van Crevel H. Measuring quality of life in stroke. Stroke 1993; 24:320-7.

4. Uyl-de-Groot C, Rutten F, Bonsel G. Measurement and valuation of quality of life in economic appraisal of cancer treatment. Eur J Cancer 1994; 30A: 111-7.

5. Nord E. The EuroQol Group: Papers written as part of the EuroQol enterprise, March 1992. In Sintonen H. (ed.) EuroQol Conference Proceedings Helsinki October 1992. Kuopio: Department of Social Sciences, University of Kuopio; 1993.

6. Nord E. The EuroQol Group: Systematic list of points made and lessons learned in the EuroQol enterprise, March 1992. In Sintonen H. (ed.) EuroQol Conference Proceedings Helsinki October 1992. Kuopio: Department of Social Sciences, University of Kuopio; 1993.

7. The EuroQol Group. EuroQol - A new facility for the measurement of health-related quality of life. Health Policy 1990; 16:199-208.

8. Krabbe P, Stouthard M, Esssink-Bot M-L, Bonsel G. The effect of adding a cognitive dimension to the EuroQol multi-attribute health status classification system. Journal of Clinical Epidemiology 1999; 52:293-301.

9. Kind P and Macran S. Levelling the playing field: increasing the number of response categories in EQ-5D. In [105].

10. van Agt H and Bonsel G. The number of levels in the descriptive system. In Busschbach Jv, Bonsel G, de Charro F. (eds.) EuroQol Plenary Meeting Rotterdam October 1993. Rotterdam: Erasmus University, 1994.

11. Selai C. Four levels and a funeral: preliminary testing of a four-level EuroQol descriptive classification system. In Badia X, Herdman M, Segura A. (eds.). EuroQol Plenary Meeting Barcelona October 1995. Barcelona: Institut Universitari de Salut Pública de Catalunya, 1996.

12. Selai C. Testing the EuroQol 3-level and 4-level descriptive classification systems. In Nord E. (ed.) EuroQol Plenary Meeting Oslo October 1996 Conference Proceedings. Oslo: National Institute of Public Health, 1997.

13. Janssen M, Birnie E, Bonsel G. Quantification of the level descriptors for the standard EQ-5D three level system and a five level version according to 2 methods. Quality of Life Research 2008;17: 463-473.

14. Lamers L. Adjustment of EQ-5D TTO valuation sets for use of an EQ-5D five level descriptive system: a pragmatic approach. In Stavem K. (ed.) 22nd Plenary Meeting of the EuroQol Group

Discussion Papers, Oslo, Norway, September 2005. Lørenskog: Helse-Øst Health Services Research Centre, 2005.

15. Cabasés J and Sanchez E. Valuation of 5-level states and back-compatibility with existing 3-level valuation sets: a Bayesian approach. In Stavem K. (ed.) 22nd Plenary Meeting of the EuroQol Group Discussion Papers, Oslo, Norway, September 2005. Lørenskog: Helse-Øst Health Services Research Centre, 2005.

16. Pickard AS, de Leon M, Kohlmann T, Cella D, Rosenbloom S. Psychometric comparison of the standard EQ-5D to a 5 level version in cancer patients. Medical Care 2007; 45: 259-63.

17. Janssen M, Birnie E, Haagsma J, Bonsel G. Comparing the standard EQ–5D three level system with a five level version. Value in Health. 2008;11:275-284.

18. Pickard AS, Kohlmann T, Janssen M, Bonsel G, Rosenbloom S, Cella D. Evaluating equivalency between response systems: application of the Rasch model to a 3-level and 5-level EQ-5D. Medical Care 2007;45:812-9.

19. Herdman M, Kind P. Chevalier J, Gudex C, de Pourvourville G. Investigation of labels for additional EQ-5D levels: Results of main study + 1. In Busschbach J, Rabin R, de Charro F. (eds.) 24th Scientific Plenary Meeting of the EuroQol Group Proceedings, Kijkduin -The Hague, the Netherlands, September 2007. Rotterdam: EuroQol Group Executive Office, 2007.

20. Lloyd A and Quadri N. Testing alternative labels for a UK English five level version of EQ-5D. In Scalone L and Mantovani L. (eds.) 25th Scientific Plenary Meeting of the EuroQol Group Proceedings, Baveno, Italy, September 2008. Rotterdam: EuroQol Group Executive Office, 2008.

21. Luo N, Li M, Liu G, Lloyd A, de Charro F, Herdman M. Developing the Chinese version of the new 5-level EQ-5D descriptive system: the response scaling approach. Quality of Life Research 2013;22:885-90.

22. Herdman M, Sanz L, Lloyd A, Badia X, Gudex, C. Qualitative testing of two new 5-level versions of the EQ-5D in Spain: Preliminary study results. In Scalone L and Mantovani L. (eds.) 25th Scientific Plenary Meeting of the EuroQol Group Proceedings, Baveno, Italy, September 2008. Rotterdam: EuroQol Group Executive Office, 2008.

23. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, Bonsel G, Badia X. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Quality of Life Research 2011; 20:1727-1736.

24. Lloyd A, Swinburn P, Edson E, Bowman L, Boye K, Janssen M, Pickard AS, de Charro F. Development of the EQ-5D-Psoriasis. In [58].

25. Arons A and Krabbe P. Considering cognition as a bolt-on dimension for the EQ-5D. In [59].

26. Yang Y, Rowen G, Brazier J, Tsuchiya, Young T, Longworth L. Testing the input of three add-ons to the EQ-5D. A pilot study. In [59].

27. Swinburn P, Edson-Heredin E, Secnik Boye K, Janssen M, Bowman L, Lloyd A. Estimating index values for the EQ-5D-Psoriasis. In [60].

28. Janssen M. The cognitive dimension revisited. A detailed study on its added value and interactions with EQ-5D core dimensions. In [61].

29. Herdman M, Fox-Rushby J, Rabin R, Badia X, Selai C. Producing other language versions of the EQ-5D. In [2].

30. Badia X, Femandez E, Segura, A. Influence of socio-demographic and health status variables on evaluation of health states in a Spanish population. European Journal of Public Health 1995; 5:87-93.

31. Fox-Rushby J and Badia X. Reviewing international language versions of the EuroQol Instrument: challenges for the future. In Badia X, Herdman M, Segura A. (eds.). EuroQol Plenary Meeting Barcelona October 1995. Barcelona: Institut Universitari de Salut Pública de Catalunya, 1996.

32. Fox-Rushby J. First steps to assessing semantic equivalence of EQ-5D: Results of a questionnaire survey to members of the EuroQol Group. In Nord E. (ed.) EuroQol Plenary Meeting Oslo October 1996 Conference Proceedings. Oslo: National Institute of Public Health, 1997.

33. Fox-Rushby J and Selai C. What concepts does the EQ-5D measure? Intentions and interpretations. In [2].

34. Rabin R, Herdman M, Fox-Rushby J, Badia X. Exploring the results of translating the EQ-5D into 11 European languages. In [2].

35. Brooks R, Coons S, De Cock P, Devlin N, De Weerdt W, Hansen P, Ikegami N, Jelsma J, Johnson J, Marra C. EQ-5D in selected countries around the world. In [2].

36. Hennessy S and Kind P. Measuring health status in children: developing and testing a child-friendly version of EQ-5D. In [105].

37. Brooks R with the EuroQol Group. EuroQol: the current state of play. Health Policy 1996; 37:53-72.

38. Devlin N and Parkin D. Guidance to users of EQ-5D value sets. In [39].

39. Szende A, Oppe M, Devlin N. (eds.) EQ-5D Value Sets: Inventory, Comparative Review and User Guide. Dordrecht: Springer, 2007.

40. Dolan P, Gudex C, Kind P, Williams A. A social tariff for EuroQol: results from a UK general population survey. Discussion Paper 138. York: Centre for Health Economics, 1995.

41. Xie F, Gaebel K, Perampaladas K, Doble B, Pullenayegum E. Comparing EQ-5D valuation studies. A systematic review and methodological reporting checklist. Med Decis Making 2014; 34:18-20.

42. Sintonen H, Weijnen T, Nieuwenhuizen M, Oppe S, Badia X, Busschbach J, Greiner W, Krabbe P, Ohinmaa A, Roset M, de Charro F. Comparison of EQ-5D VAS valuations: analysis of background variables. In [2].

43. Greiner W, Weijnen T, Nieuwenhuizen M, Oppe S, de Charro F. A European EQ-5D valuation set. In [2].

44. Busschbach J, Weijnen T, Nieuwenhuizen M, Oppe S, Badia X, Dolan P, Greiner W, Kind P, Krabbe P, Ohinmaa A, Roset M, Sintonen H, Tsuchiya A, Williams A, Yfantopoulos J, de Charro F. A comparison of EQ-5D time trade-off values obtained in Germany, the United Kingdom and Spain. In [2].

45. van Hout B, Janssen M, Feng Y, Kohlmann T, Busschbach J, Golicki D, Lloyd A, Scalone L, Kind P, Pickard S. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. Value in Health 2012;15:708-15.

46. Oppe M, Devlin N, van hout B, Krabbe P, de Charro F. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. Value in Health 2014;17: 445-63.

47. Lamers LM. The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value sets. Med Care. 2007; 45:238-44.

48. Stolk EA, Oppe M, Scalone L, Krabbe PF. Discrete Choice Modeling for the Quantification of Health States: The Case of the EQ-5D. Value in Health. 2010; 13:1005-13.

49. McCabe C, Brazier J, Gilks P, Tsuchiya A, Roberts J, O'Hagan A, Stevens K. Using rank data to estimate health state utility models. J Health Econ. 2006;25:418-31.

50. Krabbe P and Devlin N (eds.). The development of new research methods for the evaluation of EQ-5D-5L. European Journal of Health Economics 2013; 14 (Suppl 1): S1-S73.

51. Flynn TN. Using conjoint analysis and choice experiments to estimate QALY values: issues to consider. Pharmacoeconomics. 2010;28:711-22.

52. Bansback N, Brazier J, Tsuchiya A, Anis A. Using a discrete choice experiment to estimate health state utility values. J Health Econ. 2012; 31:306-18.

53. Mulhern B, Bansback N, Brazier J, Buckingham K, Cairns J, Devlin N, Dolan P, Hole AR, Kavetsos G, Longworth L, Rowen D, Tsuchiya A. Preparatory study for the revaluation of the EQ-5D tariff: methodology report. Health Technology Assessment 2014;18, 1-191.

54. Bansback N, Hole A, Mulhern B, Tsuchiya A. Further exploration of discrete choice experiments with a duration attribute to value health states. In [61].

55. Burström K, Egmar A-C, Sun S, Björk S, Lugnér A. Proceedings of the 31st Scientific Plenary Meeting of the EuroQol Group, Stockholm, Sweden. Rotterdam: Euroqol Research Foundation, 2014.

56. Mulhern B, Bansback N, Hole A, Tsuchiya A. A methodological exploration of designing DCE with duration to model EQ-5D-5L health state preferences: In [55].

57. Tsuchiya A, Bansback N, Hole A, Mulhern B. Using DCE with duration to examine the robustness of preferences across the five dimensions of the EuroQol instrument: The second paper from the FEDEV project. In [55].

58. Yfantopoulos J. (ed.) 27th Scientific Plenary Meeting of the EuroQol Group Proceedings, Athens, Greece, September, 2010. Rotterdam: EuroQol Group Executive Office, 2010.

59. Lloyd A. (ed.) Proceedings of the 28th Scientific Plenary Meeting of the EuroQol Group, Oxford September 2011. Rotterdam: EuroQol Executive Office, 2011.

60. Busschbach J. (ed.) 29th Scientific Plenary Meeting of the EuroQol Group Proceedings, Rotterdam, 2012. Rotterdam: EuroQol Group Executive Office, 2012.

61. Bryan S, Johnson J, Ohinmaa A, Xie F. 30th Scientific Plenary Meeting of the EuroQol Group, Montreal September 2013. Rotterdam: EuroQol Executive Office, 2013.

62. Angelsen V and Olsen J. QALY-league tables for a hospital. In O'Hanlon M and Buxton M. (eds.) EuroQol Plenary Meeting London October 1994 Conference Proceedings. Uxbridge: Health Economics Research Group, Brunel University, 1995.

63. Hare J, Leese B, Hardman G. Applications of EuroQol (EQ5D) in general practice: an exploratory study. In Rabin R, Busschbach Jv, de Charro F, Essink-Bot M-L, Bonsel G. (eds.) EuroQol Plenary Meeting, Rotterdam, October 1997 Discussion Papers. Rotterdam: Institute for Medical Technology Assessment, 1998.

64. Kind P, Leese B, Hardman G. Demonstrating the value of EQ-5D as a tool for monitoring health status in primary care. In Rabin R, Busschbach Jv, de Charro F, Essink-Bot M-L, Bonsel G. (eds.) EuroQol Plenary Meeting, Rotterdam, October 1997 Discussion Papers. Rotterdam: Institute for Medical Technology Assessment, 1998.

65. Stolk E and van Busschbach Jv. The performance of EuroQol in children. In Greiner W, Graf v.d. Schulenberg J-M, Piercy J. (eds.) EuroQol Plenary Meeting Hannover October 1998 Discussion Papers. Hannover; Uni-Verlag Witte,1999.

66. Roset M, Badia X, Benavides A, Herdman M. The validity of the EQ-5D in children with asthma. In Badia X, Herdman M, Roset M. (eds.) 16th Plenary Meeting of the EuroQol Group Discussion Papers, Sitges, Spain, November 1999. Barcelona; Institut Universitari de Salut Pública de Catalunya, 2000.

67. Parkin D, Rice A, Jacoby A. Use of the EQ VAS in a daily patient diary. In Badia X, Herdman M, Roset M. (eds.) 16th Plenary Meeting of the EuroQol Group Discussion Papers, Sitges, Spain, November 1999. Barcelona: Institut Universitari de Salut Pública de Catalunya, 2000.

68. Tsuchiya A, Brazier J, McColl E, Parkin D. A condition-specific instrument, a generic instrument, and a preference-based generic instrument. In [106].

69. Rebolj M and Prevolnik Rupel V. Socioeconomic inequalities in health of the Slovenian population measured by the EQ-5D instrument. In [106].

70. Leese B, Kind P, Hardman G. Measuring outcomes – managing performance: roles for EQ-5D in delivering primary health care services? In [105].

71. Szende A. The (potential) role of the EQ-5D in the international comparison of socioeconomic inequalities in health status. In [105].

72. Alli M, Ibrahim M, Lawrence D, Lockhart J, Laitner S, Brogan C. Survey on tackling health inequalities amongst older people in Brent and Harrow district of the UK: results from EQ-5D questionnaire. In [105].

73. Rebolj M, Oppe S, Oppe M, Rabin R, Szende A, Cleemput I, de Charro F, Williams A. What light does EQ-5D shed on international differences in self-reported health problems by age, sex and educational level? In [105].

74. Oppe M, Szende A, de Charro F. Development of a methodology to gain insight on the use of EQ-5D in different disease areas. In Stavem K. (ed.) 22nd Plenary Meeting of the EuroQol Group Discussion Papers, Oslo, Norway, September 2005. Lørenskog: Helse-Øst Health Services Research Centre, 2005.

75. Betegón L and Badia X. Discussion paper review of the use of the EQ-5D in cost-utility analysis. In Badia X. (ed.) 25th Scientific Plenary Meeting of the EuroQol Group, September 2006, Barcelona. Barcelona: IMS Health, 2007.

76. Pickard AS, Wilke C, Hsiang-Wen L, Lloyd A. Impact of cancer on health related quality of life: evidence using the EQ-5D. In [107].

77. Pickard AS, Wilke C, Lin H, Lloyd A. Health utilities using the EQ-5D in studies of cancer. Pharmacoeconomics 2007; 25:365-84.

78. Pickard AS, Wilke C, Jung E. Patel S, Stavem K, Lee T. Use of a preference-based measure of health (EQ-5D) in COPD and asthma. In Busschbach J, Rabin R, de Charro F. (eds.) 24th Scientific Plenary Meeting of the EuroQol Group Proceedings, Kijkduin -The Hague, the Netherlands, September 2007. Rotterdam: EuroQol Group Executive Office, 2007.

79. Pickard AS, Wilke C, Jung E, Patel S, Stavem K, Lee T. Use of a preference-based measure of health (EQ-5D) in COPD and asthma. Respiratory Medicine 2008;102:519-36.

80. Janssen M, Lubetkin E, Pickard AS. The use of the EQ-5D preference-based health status measure in adults with type 2 diabetes mellitus. In de Pouvourville G. (ed.) 26th Scientific Plenary Meeting of the EuroQol Group Proceedings, Paris, France, September 2009. Rotterdam: EuroQol Group Executive Office, 2009.

81. Pickard AS and Lin F. Is EQ-5D inferior in content to disease-specific utility measures? In [58].

82. Janssen M, Lubetkin E, Sekhobo J, Pickard AS. The use of the EQ-5D preference-based health status measure in adults with Type 2 diabetes mellitus. Diabetic Medicine 2011; 28:395-413.

83. Goldsmith KA, Dyer MT, Schofield PM, Buxton MJ, Sharples LD. Relationship between the EQ-5D index and measures of clinical outcomes in selected studies of cardiovascular interventions. Health Qual Life Outcomes 2009;7:96.

84. Dyer MT, Goldsmith KA, Sharples LS, Buxton MJ. A review of health utilities using the EQ-5D in studies of cardiovascular disease. Health Qual Life Outcomes. 2010;8:13.

85. Goldsmith KA, Dyer MT, Buxton MJ, Sharples LD. Mapping of the EQ-5D index from clinical outcome measures and demographic variables in patients with coronary heart disease. Health Qual Life Outcomes. 2010;8:54.

86. Black J, Herbison P, Lyons R, Polinder S, Derrett S. Recovery after injury: an individual patient data meta-analysis of general health status using the EQ-5D. J Trauma 2011;71:1003-10.

87. Thompson S, Holtzer-Goor K, Schaafsma R, Uyl-de Groot C. Use of a 5L EQ-5D instrument without additional labels in patients with chronic lymphocytic leukaemia. In [107].

88. Pickard AS, De Leon M, Kohlmann T, Cella D, Rosenbloom S. Psychometric comparison of the standard EQ-5D to a 5-level version in cancer patients. In [107].

89. Scalone L, Ciampichini R, Fagiuoli S, Gardini I, Gaeta L, Del Prete A, Magini G, Baldan A, Mantovani L. Comparing the standard EQ-5D-3L versus 5L version for the assessment of health of patients with liver diseases. In [58].

90. Kreimeier S and Greiner W. Patient-reported outcomes (PRO) in clinical practice in measurement of Health-Related Quality of Life (HRQoL) in patients with knee and hip replacement in German hospitals. In [59].

91. Janssen M, Busschbach J, Golicki D, Niewada M, Scalone L, Swinburn P, Pickard A. Measurement properties of the EQ-5D-5L compared to EQ-5D-3L in 8 patient groups. In [59].

92. Prevolnik Rupel V, Ogorevc M, Marusic D. The steps to evidence based decision making in Slovenian patient reported outcome measures. In [60].

93. Lin F, Samp J, Wong P, Pickard AS. Evaluating change in using patient-reported outcome measures in knee replacement: the complementary nature of the EQ-5D index and VAS scores. In [60].

94. Jalundhwala Y, Gor D, Dang J, Lin F, Pickard S. Use of the EQ-5D to predict outcomes of hip replacement surgery. In [61].

95. Lin FJ, Pickard AS, Krishnan JA *et al*. Measuring health-related quality of life in chronic obstructive pulmonary disease: properties of the EQ-5D-5L and PROMIS Short Form. In [61].

96. Szende A, Janssen B, Cabasés J. (eds.) Self-Reported Population Health: An International Perspective based on EQ-5D. Dordrecht: Springer 2014.

97. Szende A and Williams A (eds.) Measuring Self-Reported Population Health: An International Perspective based on EQ-5D. SpringMed Publishing 2004.

98. Selai C, Gaminde I, Herdman M, Johnson J, Rabin R. EQ-5D: modes of administration. In [2].

99. Gusi N, Ortega Y, García-Martin Y, Monge M. Physical activity, lifestyle and health-related quality of life of secondary students in Extremadura, Spain. In [106].

100. Klose, K and Greiner W.: Developing a cognitive dimension as a bolt-on for the EQ-5D-Y in Germany. In [59].

101. Gusi N, Pérez-Sousa M, Gozalo M, Olivares P. Validity and reliability of the Spanish EQ-5D-Y proxy version. In [61].

102. Wille N, Badia X, Bonsel G, Burström K, Cavrini G, Devlin N, Egmar A-C, Greiner W, Gusi N, Herdman M, Jelsma J, Kind P, Scalone L, Ravens-Sieberer U. Development of the EQ-5D-Y: a child-friendly version of the EQ-5D. Quality of Life Research 2010; 19:887-97.

103. Ravens-Sieberer U, Wille N, Badia X, Bonsel G, Burström K, Cavrini G, Devlin N, Egmar A-C, Gusi N, Herdman M, Jelsma J, Kind P, Olivares P, Scalone L, Greiner W. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. Quality of Life Research 2010; 19:887-97.

104. Weijnen T, Nieuwenhuizen M, Ohinmaa A, de Charro F. Construction of the EQ-net VAS and TTO databases. In [2].

105. Kind P and Macran S. (eds.) 19th Plenary Meeting of the EuroQol Group Discussion Papers, York, UK, September 2002. York: Centre for Health Economics, 2002.

106. Norinder A, Pedersen KM, Roos P. (eds.) Proceedings of the 18th Plenary Meeting of the EuroQol Group, Copenhagen, Denmark, September 2001. Lund: Swedish Institute for Health Economics, 2002.

107. Badia X. (ed.) 25th Scientific Plenary Meeting of the EuroQol Group, September 2006, Barcelona. Barcelona: IMS Health, 2007.