ELSEVIER

European Journal of Cardio-thoracic Surgery 38 (2010) 450–455

EUROPEAN JOURNAL OF CARDIO-THORACIC SURGERY

www.elsevier.com/locate/ejcts

Engineering a novel three-dimensional contractile myocardial patch with cell sheets and decellularised matrix☆,☆☆

Hiroki Hata[a,b,*], Antonia Bär[a], Suzanne Dorfman[a], Zlata Vukadinovic[a], Yoshiki Sawa[c], Axel Haverich[a,b], Andres Hilfiker[a]

[a] *Leibniz Research Laboratories for Biotechnology and Artificial Organs (LEBAO), Hannover Medical School, Hannover, Germany*
[b] *Department of Cardiac, Thoracic, Transplantation and Vascular Surgery, Hannover Medical School, Hannover, Germany*
[c] *Department of Cardiovascular Surgery, Osaka University Graduate School of Medicine, Suita, Japan*

Received 1 October 2009; received in revised form 30 January 2010; accepted 4 February 2010; Available online 23 March 2010

Abstract

Objectives: A persistent problem in generating a functional myocardial patch is maintaining contractions in a thicker construct. Thus far, we have successfully created contracting constructs with a defined directionality by seeding neonatal rat cardiomyocytes (CMs) on decellularised porcine small-intestinal submucosa (SIS). Here, we report our efforts in generating a thicker contracting construct by combining CM cell sheets with CM-seeded SIS. **Methods:** Porcine SIS was decellularised, opened along the longitudinal axis, fixed in a metal frame (45 mm × 25 mm) and seeded onto the submucosal side with neonatal rat CMs at a density of 1.8×10^5 cells cm^{-2}. CM sheets were prepared using temperature-responsive dishes by seeding CMs at a density of 4.0×10^5 cells cm^{-2}. Three days after CM seeding, one- or three-layered CMs sheet(s) were stacked onto seeded SIS. Construct contraction was observed for an additional 10 days followed by histological analysis. **Results:** Stacked CM sheets contracted spontaneously and synchronously with seeded SIS after adherence. A large portion of analysed constructs showed a defined contraction direction, parallel to the longitudinal axis (seeded SIS: 83%, seeded SIS + 1 sheet: 70%, seeded SIS + 3 layered sheets: 71%). This finding was in agreement to the histological finding of aligned CMs parallel to the longitudinal axis. The thickness of seeded SIS with and without three-layered sheets was approximately 800 μm and 500 μm, respectively. **Conclusions:** By combining layered CM sheets with CM-seeded SIS, a three-dimensional myocardial patch with contraction in a defined direction was successfully generated. This may represent an intermediate step to a multiple layered, vascularised contractile myocardial graft.

Keywords: Tissue engineering; Small-intestinal submucosa; Cell sheet; Myocardial patch

1. Introduction

A major aim of tissue engineering is the production of a functional three-dimensional structure, aimed to restore, support or supersede native tissue function. Thus far, multiple tissue-engineered constructs have been developed and clinically applied since the term 'tissue engineering' was first introduced in 1987 [1]. Above all, engineering of myocardial tissue may become a pressing need in an ageing society with increasing cardiovascular morbidity [2]. Although several reports of producing cardiac tissue have been made since the late 1950s, engineering a fully functional and transplantable heart muscle has not yet been achieved due to the structural and functional complexity of native heart muscle [1–3].

Research conducted by Zimmermann and Zhao [3,4] has suggested three potential approaches to construct functional contractile cardiac tissue: (1) seeding cardiomyocytes (CMs) on a synthetic or biological scaffold; (2) using soluble collagen and other extracellular matrix (ECM) components to entrap CMs; and (3) stacking CM sheets to form multilayered cardiac muscle constructs. With respect to approach (1), our laboratory has successfully developed a contracting artificial myocardial tissue by seeding CM on a collagen scaffold [5]. Progress in this direction was achieved through the replacement of the collagen scaffold with decellularised porcine small-intestinal submucosa (SIS) as the matrix. When seeded with neonatal rat CM, a contracting construct with a defined direction was generated [6].

Cell sheet technology (approach 3) has several advantages; for example, the preservation of cell-to-cell connec-

☆ Presented at the 23rd Annual Meeting of the European Association for Cardio-thoracic Surgery, Vienna, Austria, October 18–21, 2009.

☆☆ *Sources of funding:* This study was supported by the CORTISS Foundation. H.H. received a fellowship from the Japan Heart Foundation and the Uehara Memorial Foundation.

* Corresponding author. Address: Leibniz Research Laboratories for Biotechnology and Artificial Organs (LEBAO), Hannover Medical School, Carl-Neuberg-Strasse 1, D-30625 Hannover, Germany. Tel.: +49 511 532 8913; fax: +49 511 532 8819.

E-mail address: Hata.Hiroki@mh-hannover.de (H. Hata).

1010-7940/$ – see front matter
doi:10.1016/j.ejcts.2010.02.009

tion with ECM, the maintenance of electrophysiological and contractile performance, the feasibility of stacking multiple sheets and the independence from potentially immunogenic or pathogenic scaffold materials [1,7]. However, the maximum thickness and strength achievable, relative to native tissue, remains equivocal. Our laboratory has investigated the transfer of cell sheets of various origins, as a regenerative therapy option in different animal models [8,9] and in a clinical study in Osaka University.

Decellularised porcine SIS is an acellular ECM rich in collagen, glycosaminoglycans and growth factors, and been characterised as a complete absorbable biocompatible framework [10]. As a result of these characteristics, it has great potential for use in both *ex vivo* tissue engineering and in reconstructive surgery. SIS and similar ECM scaffolds, for example, decellularised porcine urinary bladder matrices have also been applied in animal models of cardiac repair [11,12].

Herein, we describe our next step towards an implantable myocardial patch combining our previous success with CM-seeded SIS and CM cell sheets.

2. Materials and methods

2.1. Animal care

This work was approved by the Institutional Review Board and the Institutional Animal Care and Use Committee protocols of Hannover Medical School. All animals received humane care in compliance with the European Convention on Animal Care.

2.2. Decellularisation of porcine SIS

Porcine small intestine was harvested from anaesthetised German Landrace pigs weighing 20–25 kg and cut into short segments. After mechanical removal of tunica mucosa and tunica serosa, segmented intestines were chemically decellularised using a modified method of Meezan et al. [13]. In brief, decellularisation was performed using 4% sodium-deoxycholate and 0.1% sodiumazide under continuous shaking at 4 °C for 2 h. SISs were washed with phosphate-buffered saline (PBS) containing 0.1% neomycin sulphate and 1% penicillin–streptomycin solution under continuous shaking for 7 days at 4 °C. Following the last wash, SISs were sterilised by 150 Gy gamma-ray irradiation for 75 min. Just before use, they were cut open along the longitudinal axis and fixed on a 45 mm × 25 mm metal frame.

2.3. Preparation of CM-seeded SIS and CM sheet

Neonatal rat CMs were isolated as described previously [14]. In brief, ventricles from 1- to 2-day-old Sprague Dawley rats were minced in ADS buffer (116 mM NaCl, 20 mM HEPES, 10 mM NaH_2PO_4, 5.4 mM KCl, 0.8 mM $MgSO_4$ and 5.6 mM glucose) and enzymatically digested with type II collagenase (Worthington, Lakewood, NJ, USA) and pancreatin (Sigma–Aldrich, St. Louis, MO, USA) at 37 °C. CMs were enriched in the primary CM isolates either by differential centrifugation through discontinuous Percoll gradient or with the preplating procedure. For preplating, CM isolates in culture medium were cultivated in flasks for 1 h. Following 1 h, the remaining supernatant in the flask contained an enriched CM fraction. Purification of CM by a discontinuous Percoll gradient was prepared as described before [15] using two density solutions, 1.062 and 1.082 g ml^{-1}, made from Percoll reagent (GE Healthcare). The collected cells after Percoll gradient centrifugation were seeded at a density of 1.8×10^5 cells cm^{-2} onto the submucosa side of SIS fixed in a metal frame. The SIS with cells were cultivated in 1:4 Dulbecco's Modified Eagle's Medium (Gibco, Invitrogen):medium 199 (PAA, Pasching, Austria), 10% foetal calf serum (PAA), 5% horse serum (Gibco) and 1% penicillin–streptomycin in a humidified 37 °C atmosphere with 5% CO_2.

For the generation of cell monolayer sheets, preplated CMs were cultivated in temperature-responsive dishes as described before [7]. In brief, CMs were seeded at a density of 4.0×10^6 cells per poly(N-isopropylacrylamide) grafted temperature-responsive dish (UpCell; Cellseed, Tokyo, Japan) with a diameter of 35 mm and incubated in culture medium for 3 days.

2.4. Transfer of CM sheet onto decellularised SIS

After 3 days of cultivation, CM cell sheets were detached through a temperature change from 37 °C to 20 °C for approximately 30 min. After detachment, one cell sheet ($n = 24$) or three-layered cell sheets ($n = 18$) were stacked onto a CM-seeded SIS according to previously described procedures [7]. Briefly, the entire CM sheet with media was gently aspirated and transferred onto the SIS. Media was then gently dropped onto the centre of the sheet to spread the folded parts. Excess media was aspirated to allow the cell sheet to adhere to the SIS. After 30 min and using the same protocol, additional CM sheets were transferred and spread onto the formerly attached cell sheet. Approximately 30 min were found to be sufficient for adhesion of the layered sheet(s). In addition, one CM sheet on an un-seeded SIS ($n = 13$) and CM-seeded SIS without CM sheet ($n = 12$) were prepared. All constructs were incubated at 37 °C for an additional 10 days and the culture medium was changed every 2 days.

2.5. Assessment of cell population

Cells after Percoll gradient or preplating were seeded on 10 temperature-responsive dishes (3.0×10^6 cells per dish), respectively, and incubated in a culture medium as described above for 1 day. After fixation with 4% formaldehyde, cells were stained with monoclonal anti-α-sarcomeric actinin (Sigma) or anti-prolyl-4β-hydroxylase (Acris, Hiddenhausen, Germany). Nuclei were counterstained with 4′,6-diamidino-2-phenylindole (DAPI) (Sigma). The images were photographed using an inverted research microscope (Axio Observer A1, Zeiss). Cell population percentages were calculated by dividing the number of respective cells by the total number of cells. Four randomly selected optical fields (0.083 mm^2) per dish were analysed and an image analysis software (ImageJ 1.40 g, Wayne Rasband, NIH, USA) was used.

2.6. Microscopic and macroscopic observation

All constructs were observed by means of an inverted optical microscope (CKX41, Olympus) to count the synchronous contraction rate and to observe the contraction direction. The contractile direction, within approximately 30° from the longitudinal axis of the SIS, was defined as an oriented contraction parallel to the longitudinal axis. The ratio of constructs showing oriented contractions was calculated by dividing the number of respective myocardial grafts by the total number of myocardial grafts. Microscopic images of contracting CM sheets on SIS were recorded using a motorised inverted research microscope (Axio Observer Z1, Zeiss). Macroscopic images of contracting myocardial patches were also recorded using a digital movie camera (DMX-HD2, SANYO).

2.7. Histological analysis

For assessment of collagen fibre alignment, a small piece of decellularised SIS was fixed with 4% paraformaldehyde and stained with anti-collagen I (Sigma) detected by Alexa Fluor 488 (Molecular Probes, Eugene). For characterisation of the seeded cells and cell sheets on SIS, samples were cut into approximately 4 cm^2 squares, fixed with 4% paraformaldehyde and stained with Alexa Fluor 488-conjugated phalloidin (Molecular Probes) or monoclonal anti-α-sarcomeric actinin and Alexa Fluor 488 (Molecular Probes). Cell nuclei were counterstained with DAPI. Additional samples were fixed with 10% formalin, embedded in paraffin, cut into 6-μm cross-sections, and stained with haematoxylin and eosin or monoclonal anti-connexin 43 (Sigma) and Alexa Fluor 488. Connexin 43-stained samples were counterstained with DAPI. Stained sections were analysed and documented using an inverted research microscope (Axio Observer A1).

2.8. Statistics

All data are presented as mean ± standard deviation.

3. Results

3.1. Decellularised SIS

Decellularised SIS was pliable, durable and easily stretched onto a metal frame (Fig. 1(A) and (B)). Collagen fibres in the decellularised SIS were oriented along the longitudinal axis (Fig. 1(C)).

3.2. Cell population after Percoll gradient and preplating

CMs and fibroblasts were identified through staining for α-sarcomeric actinin and prolyl-4β-hydroxylase, respectively. The cell suspension after Percoll gradient centrifugation contained 86.9 ± 3.8% CM and 10.5 ± 1.0% fibroblasts, while preplating isolates contained 74.6 ± 3.6% CM and 21.6 ± 5.6% fibroblasts.

SIS were seeded with both CM isolates; however, SIS seeded with preplated CM isolates resulted in a less synchronous contraction pattern and over a shorter duration than SIS seeded with the Percoll CM isolates (data not shown). Similarly, CM cell sheets were made with both CM isolates. Cell sheets and subsequent stacking made with Percoll CM isolates resulted in a less tight adhesion as shown by cross-section histological assessment (Fig. 2). Based on these data, CM from Percoll isolates were used as the cell source for direct seeding onto the SIS, and CM cell sheets stacked on the SIS were made with preplated CM isolates.

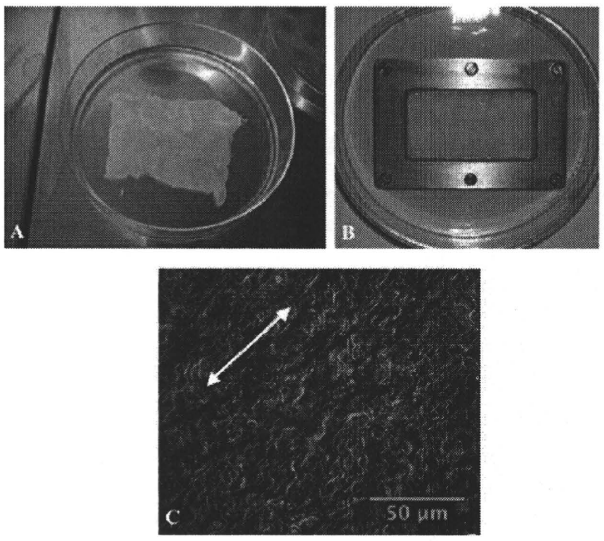


Fig. 1. Decellularised porcine small-intestinal submucosa (SIS) (A), fixed in a metal frame such that the long axis of the SIS was parallel to the long axis of the metal frame (B). Anti-collagen I staining revealed that collagen fibre in SIS aligned along the longitudinal axis (C). Scale bars: 50 μm.

3.3. Macroscopic and microscopic observation

All Percoll CM-seeded SIS started to contract spontaneously 1–2 days after initial CM seeding. After lowering culture temperature from 37 °C to 20 °C, CM cell sheets

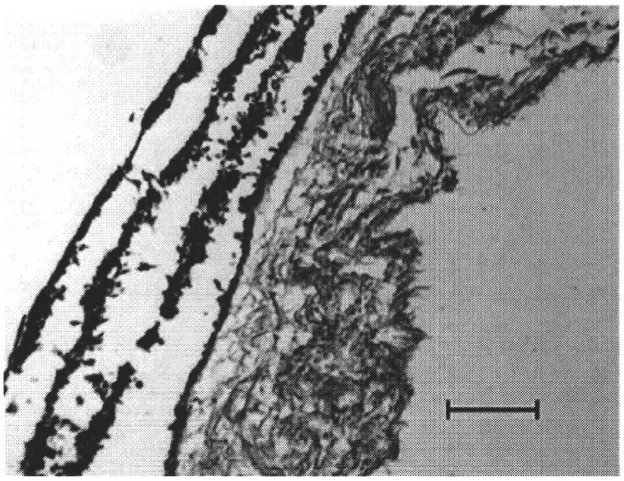


Fig. 2. Haematoxylin and eosin staining of a cross-sectional slice of a myocardial patch made with three-layered Percoll-enriched cardiomyocyte (CM) sheets on a CM-seeded small-intestinal submucosa (SIS). CM sheets did not adhere to one another or to the seeded SIS. Scale bars: 200 μm.

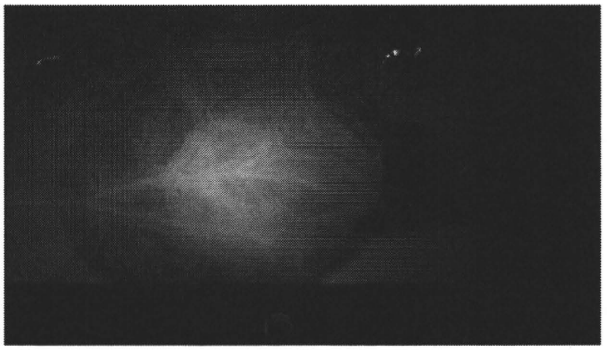


Fig. 3. Macroscopic appearance of seeded SIS with three-layered cardiomyocyte sheets.

spontaneously detached as a single unit maintaining contractility without enzymatic treatment. Initially, a transferred CM sheet shrank to some extent, and had asynchronous contractions in multiple directions. As shown in Fig. 3, the cell sheet could be manipulated to lie flat on either seeded or un-seeded SIS. After stacking, the cell sheets placed on SIS collectively contracted in a defined direction parallel to the longitudinal axis of the SIS (Supplemental Video 1). The wave-like contraction pattern was observed for both CM sheets layered on CM-seeded SIS and un-seeded SIS. All constructs, regardless of the number of cell sheets, began contracting a few minutes after transfer. Approximately 50% of all myocardial patches with one or three cell sheet(s) on CM-seeded SIS contracted until day 10 after sheet transplantation, while un-seeded SIS with one CM sheet contracted for only 4 days after cell sheet transfer. A large proportion of analysed constructs showed an oriented contraction parallel to the longitudinal axis of the SIS 3 days after cell sheet transfer; CM-seeded SIS without cell sheets: 83%, CM-seeded SIS with one CM sheet: 70%, CM-seeded SIS with three-layered CM sheets: 71%. Macroscopic observation of CM-seeded SIS with three-layered CM sheets with a defined contraction orientation and parallel to the longitudinal axis of SIS is shown in Supplemental Video 2.

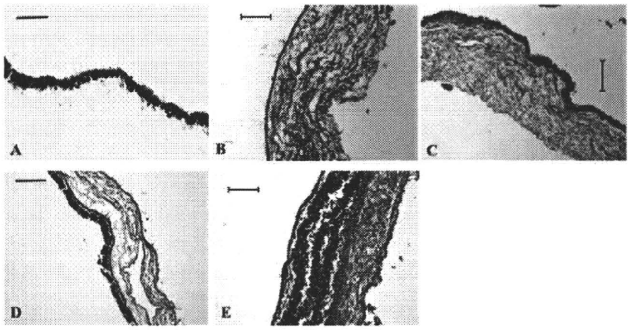


Fig. 4. Haematoxylin and eosin staining of cross-sectional slices of myocardial patches. Small-intestinal submucosa (SIS) was seeded with Percoll cardiomyocytes (CM), cell sheets were made with preplated CM. (A) single CM cell sheet; (B) decellularised porcine SIS; (C) CM-seeded SIS; (D) CM-seeded SIS with 1 CM sheet; (E) CM-seeded SIS with three-layered CM sheets. Scale bars: 100 μm (A), 200 μm (B–E).

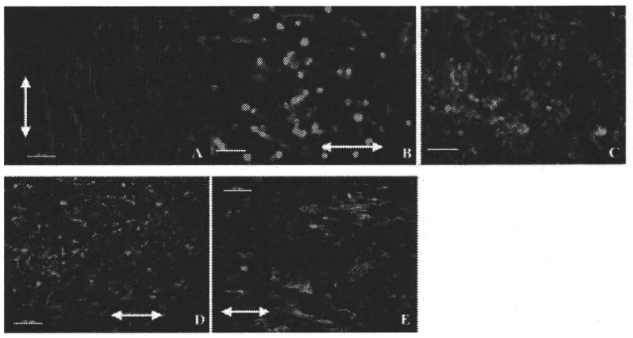


Fig. 5. Immunohistochemical assessment of decellularised small-intestinal submucosa (SIS) seeded with cardiomyocytes (CM) (A and B), CM sheet detached from the temperature-responsive dish (C), and CM sheet on seeded SIS (D and E). Cells were stained with phalloidin (A and C) or anti-α-sarcomeric actinin (B, D and E). Nuclei were stained with DAPI (A–E). Note orientation of CM cell sheet on seeded SIS (D and E) and seeded CM on SIS (A and B), which are consistent with the longitudinal axis of SIS indicated by white arrows. Scale bars: 50 μm (A, C and E), 100 μm (B and D).

3.4. Histological analysis

Cross-sectional slices stained with haematoxylin and eosin are shown in Fig. 4. As shown in Fig. 4(A) and (B) respectively, one CM cell sheet has a thickness of approximately 40–50 μm and a decellularised SIS is approximately 300–500 μm thick. No remnant cells were detected in the decellularised SIS. Seeded Percoll CM attached to the decellularised SIS (Fig. 4(C)). Layered CM cell sheets adhered to each other or to the CM-seeded SIS (Fig. 4(D) and (E)), and a seeded SIS layered with three CM sheets measured approximately 600–800 μm thick.

Immunohistochemical analysis with phalloidin and anti-α-sarcomeric actinin confirmed that the seeded Percoll CM aligned parallel to the longitudinal axis of the SIS (Fig. 5(A) and (B)). Although CM in the detached cell sheet appeared disordered and without alignment (Fig. 5(C)), once attached to the seeded SIS, the CM within the cell sheet aligned in a similar direction to the Percoll CM (along the longitudinal axis of the SIS) (Fig. 5(D) and (E)). These findings support the macroscopic and microscopic observations reported above.

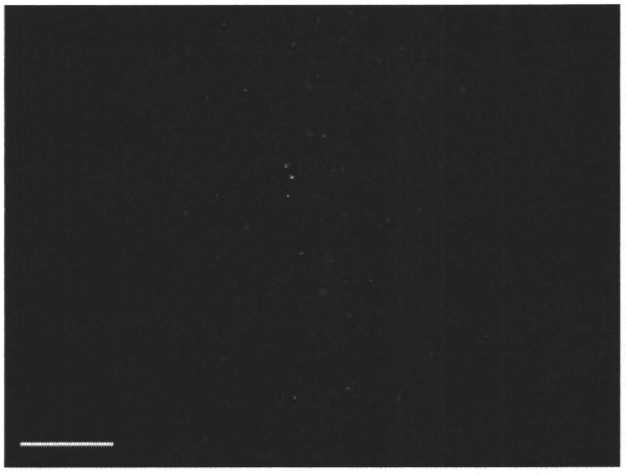


Fig. 6. Anti-connexin 43 staining revealed gap-junctions throughout three-layered cardiomyocyte sheets on seeded small-intestinal submucosa (SIS). Nuclei were counterstained with DAPI. Scale bars: 100 μm.

Anti-α-sarcomeric actinin staining also revealed the typical striated pattern of the CM in both the Percoll-seeded cells on SIS and in the CMs of the attached sheets (Fig. 5(B), (D) and (E)). Immunohistochemical staining with anti-connexin 43 demonstrated a diffuse expression of connexin 43 within the three-layered CM sheets on the seeded SIS (Fig. 6).

4. Discussion

Cell-transplantation therapy, to treat damaged myocardium, has been widely studied and has resulted in positive, albeit limited, effects in clinical trials [16]. Despite early success, problems still remain including method of cell delivery, poor graft survival, functional integration and post-transplantation arrhythmias [17]. Moreover, changes in the ECM, which are regarded to play a pivotal role for cell survival, differentiation, proliferation, metabolism and integrative function, are not taken into consideration with this intervention [18]. The emergence of a second regenerative therapy using a combination of cells and ECM provide greater advantages towards the engineering of a three-dimensional functional myocardial patch [18,19]. To this point, the present study demonstrated the feasibility of generating a contractile myocardial patch by combining a decellularised biological matrix (SIS) with CM cell sheets. Moreover, we show that CM sheets can be layered on SIS without disrupting contractile alignment and function of the CM. Thus, our results represent an important intermediate step towards the engineering of a complex, multilayered contractile myocardial patch.

Decellularised SIS is an ideal matrix for tissue engineering and remodelling due in part, to its intact ECM. The ECM of SIS consists primarily of collagens (types I, III and VI), and also glycosaminoglycans, for example, hyaluronic acid, chondroitin sulphate A and B, heparin, and heparan sulphate, and glycoproteins, for example, fibronectin. In addition, several cytokines are produced, including basic fibroblast growth factor and transforming growth factor-β, which together induce cellular migration, proliferation and differentiation, early capillary ingrowth and ultimately, endothelialisation. In addition, decellularised SIS has been previously characterised as a complete absorbable biocompatible framework with a high resistance towards infection [20,21]. Moreover, inherent SIS fibre alignment is along the longitudinal axis of the small intestine [22,23], which enables us to induce a 'preferred' cellular alignment without mechanical stimulation. Typically, the native cell alignment in an artificial tissue-engineered system is accomplished either by mechanical stimulation or by using a patterned scaffold requiring complex devices or materials [1,4]. When SIS is used; however, it leads to an orthotropic mechanical behaviour of the scaffold, with the preferred fibre direction showing great stiffness and strength [22,23].

Cell sheet engineering is a well-established technique using cells from various sources, and has already been investigated in a clinical setting [24]. CM cell sheets are unique due to their contractile ability. The most distinct advantage of CM cell sheet engineering is that they can be easily layered while maintaining electrically communicative pulsation supported through gap-junctions [7,8]. To this point, anti-connexin 43 immunohistochemistry revealed diffuse gap-junctions within the three-layered CM sheets on the CM-seeded SIS. This abundant detection of connexin 43 suggested the formation of an electrical syncytium throughout the cell sheets and connection with CM seeded on SIS. This finding, in addition to the intact ECM of the SIS, provides strong support for the coupling of these two technologies.

Despite initial success, additional questions with respect to the detailed character of a tissue-engineered heart patch remain unanswered. First, the strength and durability of the contractile forces generated by the layered CM sheets are unknown. These assessments are crucial as the patch is being generated to replace scar tissue in the left ventricle of the heart. A previous study has shown that SIS itself or a similar ECM, for example, porcine urinary bladder matrix, could substitute for the right ventricular wall [11] or even the left ventricular wall [12] in a large animal model. However, in that study, the patches were implanted as four ECM-layered patches based on the idea that the mechanical behaviour of a single layer of SIS is insufficient for most load-bearing applications [22]. Recently, our colleagues reported a successful implantation of an autologous vascularised matrix from a porcine small bowel segment without mucosa to the right ventricular wall of the pig [25]. While it differs from our myocardial construct in its characteristics, its thickness is comparable. The results from this study suggest that one SIS layer may, in fact, be sufficient.

A second question is whether the ECM provided by the SIS is similar or adaptable to a native-like ECM, to support the contractility of the ventricle [25]. While the CM cell sheet has an ECM of ventricular origin, the effect of the intestinal ECM on myocardial function has not been determined. Cell source is the subject of the third question. For clinical application, the use of autologous, allogenic or xenogenic CMs may be problematic. Of course, this matter is universal in cardiac tissue engineering and most cell-based therapies. To this point, embryonic and adult stem cells, or induced pluripotent stem cell-based approaches will play a central role in the near future [1].

Finally, the engineering of myocardial heart patch with CM cell sheets is limited by the thickness generated through sheet stacking. In this study, a myocardial patch with a thickness of roughly 800 μm was built. Although possible, more cell sheets can be continuously layered; however, supplying sufficient oxygen and nutrients will challenge the viability of a thicker construct. Although these issues will be addressed to determine the future clinical application, the present preliminary study shows the novel development of a pulsatile, thicker myocardial patch with unidirectional contractions and proposes a new strategy for regenerative therapy for cardiac dysfunction.

In conclusion, by combining a CM-seeded matrix with CM sheets, a multilayered and contracting myocardial graft was successfully generated. Moreover, our novel myocardial patch is a good foundation for our ultimate goal; a vascularised three-dimensional myocardial graft.

References

[1] Eschenhagen T, Zimmermann WH. Engineering myocardial tissue. Circ Res 2005;97:1220–31.

[2] Haverich A. Cardiac tissue engineering. Eur J Cardiothorac Surg 2008;34:227–8.
[3] Zimmermann WH, Didie M, Döker S, Melnzchenko I, Naito H, Rogge C, Tiburcz M, Eschenhagen T. Heart muscle engineering: an update on cardiac muscle replacement therapy. Cardiovasc Res 2006;71:419–29.
[4] Zhao YS, Wang CY, Li DX, Zhang XZ, Qiao Y, Guo XM, Wang XL, Dun CM, Dong LZ, Song Y. Construction of a unidirectionally beating 3-dimensional cardiac muscle construct. J Heart Lung Transplant 2005;24:1091–7.
[5] Kofidis T, Akhyari P, Boublik J, Theodorou P, Martin U, Ruhparwar A, Fischer S, Eschenhagen T, Kubis HP, Kraft T, Leyh R, Haverich A. In vitro engineering of heart muscle: artificial myocardial tissue. J Thorac Cardiovasc Surg 2002;124:63–9.
[6] Bär A, Haverich A, Hilfiker A. Cardiac tissue engineering: "reconstructing the motor of life". Scand J Surg 2007;96:154–8.
[7] Shimizu T, Yamato M, Isoi Y, Akutsu T, Setomaru T, Abe K, Kikuchi A, Umezu M, Okano T. Fabrication of pulsatile cardiac tissue grafts using a novel 3-dimensional cell sheet manipulation technique and temperature-responsive cell culture surfaces. Circ Res 2002;90:e40.
[8] Miyagawa S, Sawa Y, Sakakida S, Taketani S, Kondoh H, Memon IA, Imanishi Y, Shimizu T, Okano T, Matsuda H. Tissue cardiomyoplasty using bioengineered contractile cardiomyocyte sheets to repair damaged myocardium: their integration with recipient myocardium. Transplantation 2005;80:1586–95.
[9] Hata H, Matsumiya G, Miyagawa S, Kondoh H, Kawaguchi N, Matsuura N, Shimizu T, Okano T, Matsuda H, Sawa Y. Grafted skeletal myoblast sheets attenuate myocardial remodeling in pacing-induced canine heart failure model. J Thorac Cardiovasc Surg 2006;32:918–24.
[10] Hodde J. Naturally occurring scaffolds for soft tissue repair and regeneration. Tissue Eng 2002;8:295–308.
[11] Badylak S, Obermiller J, Geddes L, Matheny R. Extracellular matrix for myocardial repair. Heart Surg Forum 2003;6:E20–26.
[12] Robinson KA, Li J, Mathison M, Redkar A, Cui J, Chronos NA, Matheny RG, Badylak SF. Extracellular matrix scaffold for cardiac repair. Circulation 2005;112:I135–143.
[13] Meezan E, Hjelle JT, Brendel K. A simple versatile; nondisruptive method for the isolation of morphologically and chemically pure basement membranes from several tissues. Life Sci 1975;17:1721–32.
[14] Wollert KC, Taga T, Saito M, Narazaki M, Kishimoto T, Glembotski CC, Vernallis AB, Heath JK, Pennica D, Wood WI, Chien KR. Cardiotrophin-1 activates a distinct form of cardiac muscle cell hypertrophy. Assembly of sarcomeric units in series VIA gp130/leukemia inhibitory factor receptor-dependent pathways. J Biol Chem 1996;271:9535–45.
[15] Iwaki K, Sukhatme VP, Shubeita HE, Chien KR. Alpha- and beta-adrenergic stimulation induces distinct patterns of immediate early gene expression in neonatal rat myocardial cells. fos/jun expression is associated with sarcomere assembly; Egr-1 induction is primarily an alpha 1-mediated response. J Biol Chem 1990;265:13809–17.
[16] Hagège AA, Marolleau JP, Vilquin JT, Alhéritière A, Peyrard S, Duboc D, Abergel E, Messas E, Mousseaux E, Schwartz K, Desnos M, Menasché P. Skeletal myoblast transplantation in ischemic heart failure: long-term follow-up of the first phase I cohort of patients. Circulation 2006;114(1 Suppl.):I108–13.
[17] Menasché P, Alfieri O, Janssens S, McKenna W, Reichenspurner H, Trinquart L, Vilquin JT, Marolleau JP, Seymour B, Larghero J, Lake S, Chatellier G, Solomon S, Desnos M, Hagège AA. Circulation 2008;117:1189–200.
[18] Akhyari P, Kamiya H, Haverich A, Karck M, Lichtenberg A. Myocardial tissue engineering: the extracellular matrix. Eur J Cardiothorac Surg 2008;34:229–41.
[19] Chachques JC, Trainini JC, Lago N, Cortes-Morichetti M, Schussler O, Carpentier A. Myocardial Assistance by Grafting a New Bioartificial Upgraded Myocardium (MAGNUM trial): clinical feasibility study. Ann Thorac Surg 2008;85:901–8.
[20] Lindberg K, Badylak SF. Porcine small intestinal submucosa (SIS): a bioscaffold supporting in vitro primary human epidermal cell differentiation and synthesis of basement membrane proteins. Burns 2001;27:254–66.
[21] Mertsching H, Walles T, Hofmann M, Schanz J, Knapp WH. Engineering of a vascularized scaffold for artificial tissue and organ generation. Biomaterials 2005;26:6610–7.
[22] Badylak SF. The extracellular matrix as a biologic scaffold material. Biomaterials 2007;28:3587–93.
[23] Sacks MS, Gloeckner DC. Quantification of the fiber architecture and biaxial mechanical behavior of porcine intestinal submucosa. J Biomed Mater Res 1999;46:1–10.
[24] Yang J, Yamato M, Shimizu T, Sekine H, Ohashi K, Kanzaki M, Ohki T, Nishida K, Okano T. Reconstruction of functional tissues with cell sheet engineering. Biomaterials 2007;28:5033–43.
[25] Tudorache I, Kostin S, Meyer T, Teebken O, Bara C, Hilfiker A, Haverich A, Cebotari S. Viable vascularized autologous patch for transmural myocardial reconstruction. Eur J Cardiothorac Surg 2009;36:306–11.

Appendix A. Conference discussion

Dr P. Menasche *(Paris, France)*: Thank you and your colleagues for providing additional evidence that the expected benefits of cell therapy are likely to be enhanced by some form of cell scaffolding. That said, I have two questions.

The first is that if one looks at the improvement in the contraction patterns in your composite biomaterial, the composite – I mean the SIS plus the cell sheets, it turns out that it might simply reflect the fact that you had many more cells in your combined biomaterial compared with either treatment alone. You had five times more cells in the combined scaffold. So keeping in mind that the widespread clinical application of any technique requires it to be as simple as possible, my naive question is: Would it be conceivable that you would get a similar result if you simply increased the cell-seeding density on your SIS to match the final number of cells that you have got when you have combined the two materials? And that would probably be true also for explaining the increased thickness. Again you had such a disparity in the numbers of cells between the groups that it might be one explanation for your results. So is there any rationale for combining the two scaffolds? And you probably should tell us what is the contribution of a given material to the other instead of using just a single material with more cells, contractile cells, on board.

My second question is that you have shown us that there was some expression of connexin 43. Did you do some dye experiments documenting that you have an effective transmission of the impulses from one cell to the other? And more importantly, we realize that this is an in vitro study, but do you have any in vivo data allowing you to speculate on how these epicardially-delivered cells would couple with the underlying host cardiomyocytes, because obviously this is critical.

Dr Hata: As to your first question, we seeded about 180,000 cells/sq cm onto SIS and 400,000 cells/cm sq/cardiomyocyte sheet. So, as you mentioned, these cell numbers do not match. I think we need more cells to generate cell sheet than to seed on SIS. We seeded cardiomyocyte on SIS for the sake of the affinity for cardiomyocyte sheets, and we tested also cardiomyocyte sheet on un-seeded SIS but that construct did not contract longer than 4 days in a defined direction.

And the reason we used SIS as a scaffold is not only to gain thickness, but also looking toward the next step in our research utilizing vascularised SIS, so-called BioVaM. We would like to use such vascularised SIS as our next step, so we tested the feasibility of SIS in combination with cell sheets.

And as for the second question, in fact we didn't do any functional testing or *in vivo* studies. I think it may be the focus of our future studies. However, when we think about the animal model, we used here neonatal rat cardiomyocytes. So when we use a large animal model, we must think about a suitable cell source, such as iPS cells.

And you mentioned connexin 43 – I'm sorry what was your question?

Dr Menasche: The question is: Do you have any data, or can you speculate, on how cells which are delivered on the epicardium, which would be delivered on the epicardium, how you think they could couple with the cardiomyocytes of the recipient heart which are underneath?

Dr Vaage *(Oslo, Norway)*: Does Axel Haverich want to comment on this?

Dr Haverich: We did implant, also in large animal preparations, the SIS only, and we have seen coupling there. We only did these experiments to increase the thickness. Because we can do right atrial, right ventricular replacement at this point, but for left ventricular replacement we need to increase the thickness and this is why we have this combined approach. We have done electrophysiology and put a pacemaker lead on one side of the graft *in vitro* and have seen propagation of the impulse. With regard to your connexin question, we have seen propagation of the contraction towards the other end of the graft. So there is electrophysiological contact between the cells.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejcts.2010.02.009.

Research Article

Statistics in Medicine

Received 1 February 2010, Accepted 15 July 2010 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/sim.4054

Posterior maximization and averaging for Bayesian working model choice in the continual reassessment method

T. Daimon,[a,b*†] S. Zohar[c] and J. O'Quigley[d]

The continual reassessment method (CRM) is a method for estimating the maximum tolerated dose in a dose-finding study. Traditionally, use is made of a single working model or 'skeleton' idealizing an underlying true dose–toxicity relationship. This working model is chosen either by discussion with investigators or published data, before the beginning of the trial or simply on the basis of operating characteristics. To overcome the arbitrariness of the choice of such a single working model, Yin and Yuan (*J. Am. Statist. Assoc.* 2009; 104:954–968) propose a model averaging over a set of working models. Here, instead of averaging, we investigate some alternative Bayesian model criteria that maximize the posterior distribution. We propose three adaptive model-selecting CRMs using the Bayesian model selection criteria, in which we specify in advance a collection of candidate working models for the dose–toxicity relationship, especially initial guesses of toxicity probabilities, and adaptively select the only one working model among the candidates updated by using the original CRM for each working model, based on the posterior model probability, the posterior predictive loss or the deviance information criteria, during the course of the trial. These approaches were compared via a simulation study with the model averaging approach. Copyright © 2011 John Wiley & Sons, Ltd.

Keywords: dose-finding; phase I; posterior model probability; posterior predictive loss; deviance information criterion

1. Introduction

The purpose of a dose-finding clinical trial is to determine the maximum tolerated dose (MTD) of a new agent or combination of drugs for subsequent use in phase II and phase III trials. For cytotoxic agents, the MTD is in practice a dose with some given acceptable rate of toxicity, and is determined through a sequential allocation rule.

A number of designs to determine the MTD have been developed for phase I dose-finding trials. Nonparametric designs such as the standard '3+3' design [1], its extension '$A+B$' designs [2], grouped up-and-down designs [3, 4] are easy to understand and implement since they do not require explicit model specification for a dose–toxicity relationship. On the other hand, innovative designs have been proposed such as the continual reassessment method (CRM) [5]. These are often referred to as model-based approaches [6, 7], since a single-parameter working model (sometimes called a skeleton or an initial guess for the dose–toxicity relationship) needs to be specified prior to the trial beginning. In practice, only one working model is usually chosen.

For the CRM there are a wide number of choices of model family and the particular parameterization for any model. Shen and O'Quigley [8] provided sufficient conditions for a broad range of models to correctly locate the MTD for large samples. Simulations, under an extensive array of possible situations, have shown that the CRM, along with other recently developed techniques, outperforms the standard '3+3' designs [9, 10]. Model-based approaches, such as the CRM, use fewer patients overall, reach the MTD more quickly using fewer patients and treat more patients at and close to the MTD.

[a]*Division of Biostatistics, Hyogo College of Medicine, Hyogo, Japan*
[b]*Medical Center for Translational Research, Osaka University Hospital, Osaka, Japan*
[c]*Inserm, U717, Biostatistics Department, F75010 Paris, France*
[d]*Inserm, Laboratoire de Statistiques Théorique et Appliquée, Université Paris 6, F75005 Paris, France*
**Correspondence to: T. Daimon, Division of Biostatistics, Hyogo College of Medicine, 1-1 Mukogawa-cho, Nishinomiya, Hyogo 663-8501, Japan.*
†*E-mail: daimon@hyo-med.ac.jp*

In various methodological developments for the CRM, its modified versions and applications, an important issue is the requirement of pre-specification of the working model to be used in the trial. As the toxicity profile is often unknown for a new drug, such pre-specification is largely arbitrary. O'Quigley and Zohar [11] have pointed out the risk in choosing a 'reasonable' class of working models. A 'reasonable' working model is one that would exhibit good robustness properties. Some working models, while respecting the constraints of Shen and O'Quigley [8], could be anticipated not to be reasonable in this sense. When the algorithm attempts to distinguish competing dose levels as candidates for the current estimate of the MTD, the more the spread out, the better the dose levels. As an example, if we consider three dose levels, two of which are very close while the third is well removed from both of them, then it is easier deciding in favor of this third dose as opposed to choosing between the other two. If we are given I, the total number of possible doses then it is usually an idea to spread them as much as we can, specifically, in the case where the doses lie in the interval (0,1), to roughly divide this interval into $I+1$ subintervals of a similar length. When using the original CRM proposed by O'Quigley *et al.* [5], the underparameterized working model is usually chosen for the dose–toxicity model, combined with the vague prior density. Overall fit is not an important objective of phase I studies [5, 12] and for this reason such a one-parameter working model is rich enough for the purpose of identification of the MTD.

Some authors have proposed methods either calibrating the working model prior to the beginning of the trial [13, 14] or while the trial is being carried out [15]. Cheung and Chappell [13] have illustrated that operating characteristics are less sensitive to some working model choices than others. These authors have developed a numerical technique to evaluate the model sensitivity in the CRM in order to choose one working model that will be used in the clinical trial. Lee and Cheung [14] have recently used indifference intervals, and the working model used in the CRM is then selected by specifying a range of acceptable toxicity probabilities in addition to the target probability of toxicity. An algorithm was proposed for obtaining the indifference interval that maximizes the average percentage of correct selection across a set of scenarios of true probabilities of toxicity and providing a systematic approach for selecting a working model before the beginning of the trial. A different methodological approach is that of Yin and Yuan [15]. They proposed a Bayesian model averaging (BMA) approach to obtain the posterior estimates for the true toxicity probabilities by weighing the estimates from each working model with the corresponding posterior model probability (PMP).

Our aim is to investigate model selection in the CRM in the spirit of Yin and Yuan [15]. The proposed methods can include all the elicited or possible working models provided by investigators. In this case, there is no need to choose only one working model prior to the beginning of the trial or to average several working models into a single one during the course of the trial. Only one working model is adaptively and sequentially selected during the course of the trial, based on the Bayesian model selection criteria. The model selection criteria used in our approach are the posterior predictive loss (PPL) [16], the deviance information criterion (DIC) [17] and the PMP [15, 18].

In Section 2 the original CRM is briefly described. In Section 3 Bayesian model selection criteria are introduced, and three adaptive model-selecting CRMs are proposed. In Section 4 we compare these approaches with the original CRM and the CRM using the Bayesian model averaging [15], through a simulation study with several scenarios.

2. Basic method

The basic structure of the CRM was described by O'Quigley *et al.* [5], where the details can be found. Here we recall the main ideas.

Let n be the number of patients included in the trial on I fixed ordered doses, $d_1, \ldots, d_I$. During the trial, a pair of two random variables (X_j, Y_j) is observed for the jth patient ($j=1, \ldots, n$), where X_j is a dose administered to the jth included patient, which takes a real value $x_j \in \{d_1, \ldots, d_I\}$, and Y_j is a binary response variable for toxicity after the patient is administered dose x_j, where it takes one for a toxic response and zero otherwise.

The probability of suffering from toxicity, for the jth patient, at dose x_j, is given by

$$R(x_j)=\Pr(Y_j=1|X_j=x_j)\approx\psi(x_j,a), \quad j=1,\ldots,n,$$

where $R(x_j)$ is the true probability of suffering from a toxicity and $\psi(x_j,a)$ only provides a working approximation to $R(x_j)$. We restrict our attention to a simple working model [12], given by $\psi(d_i,a)=\alpha_i^a$; $i=1,\ldots,I$. A single value of α_i for dose level i is usually pre-specified by investigators such that $\alpha_i<\alpha_{i+1}$ for $i=1,\ldots,I-1$, and a ($0<a<\infty$) is a parameter to be estimated. After the first j included patients' data $\Omega_j=\{(x_l, y_l); l=1,\ldots,j\}$ ($j=1,\ldots,n$) (pairs of dose and response) are obtained, the posterior density for parameter a is given by

$$f(a|\Omega_j)=\frac{L(\Omega_j)g(a)}{\int_0^\infty L(\Omega_j)g(a)\,\mathrm{d}a},$$

where $L(\Omega_j)$ is the likelihood, given by $L(\Omega_j)=\prod_{l=1}^{j}\{\psi(x_l,a)\}^{y_l}\{1-\psi(x_l,a)\}^{(1-y_l)}$, and $g(a)$ represents a vague prior. In O'Quigley *et al.* [5] a gamma prior is suggested and, mostly, it is enough to consider the special case of the gamma $g(a)=f(a|\Omega_0)=\exp(-a)$, that is, we assume a unit exponential distribution for the prior of a.

If a is estimated by $\tilde{a}_j=\int_0^\infty af(a|\Omega_j)\mathrm{d}a$ after the observation of the jth patient, the estimated probability of toxicity for patient j treated at dose $X_j=x_j$ is given by $\tilde{R}(x_j)=\psi(x_j,\tilde{a}_j)$. The CRM allocates the dose level with the estimated toxicity probability closest to any target θ^* (from 0 to 1), to the $(j+1)$th patient, and hence estimates the MTD as the dose level at which $(n+1)$th patient would be treated.

3. Model-selecting continual reassessment methods

To overcome the arbitrariness in pre-specification of a single working model, especially for a phase I trial in which initial guess of the toxicity probabilities is rarely accurate, as well as to avoid poor pre-specification, our proposal consists of the following procedures: (1) to use all elicited or possible working models corresponding to initial guesses of the toxicity probabilities given by investigators before the start of the trial, (2) to update each of them by the CRM simultaneously, (3) to select one working model out of them, automatically and adaptively by using some criterion, during the course of the trial, and (4) to estimate the MTD based on the selected working model and allocate the estimated MTD to each included patient.

In the following, three Bayesian model selection criteria: the PPL [16], the DIC [17] and the PMP [15, 18], are proposed and adapted to the sequential feature of dose-finding designs, and a Bayesian model averaging CRM (CRM-BMA), which has been recently proposed by Yin and Yuan [15] is described. A dose-finding algorithm of the model-adaptive CRM using either of the above three criteria are given at the end of the section.

3.1. Posterior model probability

Suppose that there exists a set of M working models corresponding to initial guesses of I toxicity probabilities for the available doses, $d_1,\ldots,d_I$, which reflects investigators' quite different opinions, before the start of the trial, on the underlying true probability of toxicity associated with each of the doses. Let us denote these as $\{(\alpha_{1,1},\ldots,\alpha_{I,1}),\ldots,(\alpha_{1,M},\ldots,\alpha_{I,M})$. Then the mth working model used in the CRM is given by

$$\psi(d_i,a_m)=\alpha_{i,m}^{a_m},$$

where a_m is the parameter for the mth working model. In the same way as the original CRM (see Section 2), after the first j included patients' data Ω_j $(j=1,\ldots,n)$ are obtained, the posterior density for the parameter a_m is given by

$$f(a_m|\Omega_j)=\frac{L_m(\Omega_j)g(a_m)}{\int_0^\infty L_m(\Omega_j)g(a_m)\mathrm{d}a_m},$$

where $L_m(\Omega_j)$ is the likelihood for the mth working model, given by $L_m(\Omega_j)=\prod_{l=1}^{j}\{\psi(x_l,a_m)\}^{y_l}\{1-\psi(x_l,a_m)\}^{(1-y_l)}$, and $g(a_m)=f(a_m|\Omega_0)=\exp(-a_m)$.

Under the mth working model a Bayesian estimator of the toxicity probability, for the jth included patient, at the ith dose d_i, is given by $\theta_{ij,m}=\psi(d_i,\tilde{a}_{j,m})$, where $\tilde{a}_{j,m}=\int_0^\infty a_m f(a_m|\Omega_j)\mathrm{d}a_m$. For the mth working model we hold $d_{i^*,m}$ such that

$$i^*=\underset{i\in(1,\ldots,I)}{\operatorname{argmin}}\ |\theta_{ij,m}-\theta^*|.$$

Here, let us consider selecting only one working model and allocating the dose level found on it to $(j+1)$ th included patient. A simple idea for selecting one working model out of a set of M working models would be to use the PMP. That is, we can select the working model with the highest PMP value. This method was briefly mentioned by Yin and Yuan [15], where the normal prior was used, and the estimation method for the toxicity probabilities was different from ours. Denoting the mth working model as M_m, the PMP for the mth working model is defined by

$$\Pr(\mathsf{M}_m|\Omega_j)=\frac{\Pr(\mathsf{M}_m)\int_0^\infty L_m(\Omega_j)f(a_m|\Omega_j)\mathrm{d}a_m}{\sum_{m=1}^{M}\Pr(\mathsf{M}_m)\int_0^\infty L_m(\Omega_j)f(a_m|\Omega_j)\mathrm{d}a_m},\tag{1}$$

where $\Pr(\mathsf{M}_m)$ is such that $\sum_{m=1}^{M}\Pr(\mathsf{M}_m)=1$ and represents the probability that the mth working model is true. $\Pr(\mathsf{M}_m)=1/M$ indicates the probability for each model and in this case the PMP results in the ratio of the marginal likelihoods. We call the CRM using this PMP criterion as the CRM-PMP.

In this connection, the BMA estimate for the toxicity probability, for the jth included patient, at dose level i, under model M_m, can be derived from the PMP, and be applied to find the dose (see [15] for the details):

$$\bar{\theta}_{ij} = \sum_{m=1}^{M} \psi(d_i, \tilde{a}_{j,m}) \Pr(\mathsf{M}_m | \Omega_j).$$

Thus, we then find the dose $d_{i^*_{\mathrm{BMA}}}$ such that

$$i^*_{\mathrm{BMA}} = \underset{i \in (1,\ldots,I)}{\operatorname{argmin}} \; |\bar{\theta}_{ij} - \theta^*|.$$

The BMA approach automatically assigns a higher weight to a better-fitting model. In Section 4 we will use it for comparisons with other approaches.

3.2. *Posterior predictive loss*

With prediction in mind, we can think of the unknown as a future observation which is a replicate of the jth patient's toxicity data y_j. If we denote it by $y^{\mathrm{rep}}_{j,m}$ for the mth working model and assume y_j and $y^{\mathrm{rep}}_{j,m}$ have the same distribution, we can define a modification of the posterior predictive loss criterion that was proposed by Gelfand and Ghosh [16], as follows:

$$PPL_{j,m} = \min_{\mathscr{A}_{j,m}} \mathrm{E}_{y^{\mathrm{rep}}_{j,m} | \Omega_j, m} [\Delta(y^{\mathrm{rep}}_{j,m}, \mathscr{A}_{j,m}; \Omega_j, \tilde{a}_{j,m})], \tag{2}$$

where the expectation in equation (2) is with respect to the posterior predictive distribution associated with $y^{\mathrm{rep}}_{j,m}$, given by,

$$p(y^{\mathrm{rep}}_{j,m} | \Omega_j) = \int_0^\infty p(y^{\mathrm{rep}}_{j,m} | a_m) f(a_m | \Omega_j) \mathrm{d}a_m,$$

where $p(y^{\mathrm{rep}}_{j,m} | a_m)$ is assumed to be a Bernoulli distribution with probability $\alpha^{a_m}_{i,m}$ of $y^{\mathrm{rep}}_{j,m} = 1$, denoted by $\mathrm{Bern}[\alpha^{a_m}_{i,m}]$ and $\mathscr{A}_{j,m}$ is an action for the jth included patient under the mth working model, trying to accommodate both y_j and what we predict for $y^{\mathrm{rep}}_{j,m}$. Here, for $y^{\mathrm{rep}}_{j,m}$ and $\mathscr{A}_{j,m}$, with a univariate loss function $\Delta(y, \mathscr{A})$, we can define

$$\Delta(y^{\mathrm{rep}}_{j,m}, \mathscr{A}_{j,m}; \Omega_j, \tilde{a}_{j,m}) = \Delta(y^{\mathrm{rep}}_{j,m}, \mathscr{A}_{j,m}) + q\Delta(y_j, \mathscr{A}_{j,m}), \quad q \geqslant 0,$$

where the specified weight q indicates the relative regret for departure from y_j compared with departure from $y^{\mathrm{rep}}_{j,m}$. In particular, the case when $q = 0$ represents that the action $\mathscr{A}_{j,m}$ is a 'guess' for $y^{\mathrm{rep}}_{j,m}$. If we take $\mathscr{A}_{j,m}$ as y_j to avoid the arbitrariness of choice of q, we no longer require minimization over $\mathscr{A}_{j,m}$ in equation (2), and further can define the loss function using the logarithm of the product of the likelihood ratio for each patient:

$$\Delta(y^{\mathrm{rep}}_{j,m}, \mathscr{A}_{j,m}; \Omega_j, \tilde{a}_{j,m}) = \Delta(y^{\mathrm{rep}}_{j,m}, y_j) = \log \prod_{l=1}^{j} \frac{\{\psi(x_l, \tilde{a}_{l,m})\}^{y^{\mathrm{rep}}_{l,m}} \{1 - \psi(x_l, \tilde{a}_{l,m})\}^{(1 - y^{\mathrm{rep}}_{l,m})}}{\{\psi(x_l, \tilde{a}_{l,m})\}^{y_l} \{1 - \psi(x_l, \tilde{a}_{l,m})\}^{(1 - y_l)}}.$$

Therefore we can redefine $\mathrm{PPL}_{j,m}$ in equation (2) as:

$$\mathrm{PPL}_{j,m} = \mathrm{E}_{y^{\mathrm{rep}}_{j,m} | \Omega_j, m} \left[\log \prod_{l=1}^{j} \frac{\{\psi(x_l, \tilde{a}_{l,m})\}^{y^{\mathrm{rep}}_{l,m}} \{1 - \psi(x_l, \tilde{a}_{l,m})\}^{(1 - y^{\mathrm{rep}}_{l,m})}}{\{\psi(x_l, \tilde{a}_{l,m})\}^{y_l} \{1 - \psi(x_l, \tilde{a}_{l,m})\}^{(1 - y_l)}} \right]. \tag{3}$$

The expectation over $y^{\mathrm{rep}}_{j,m}$ can be calculated using the Monte Carlo integration. Thus, we can select the working model with the smallest value of the PPL in equation (3). We call the CRM based on the PPL criterion as the CRM-PPL.

3.3. *Deviance information criterion*

The deviance information criterion for the jth included patient under the mth working model is defined by

$$\mathrm{DIC}_{j,m} = 2\bar{D}_{j,m} - D_{j,m}, \tag{4}$$

where $\bar{D}_{j,m}$ is the posterior mean of the log-likelihood, given by $\bar{D}_{j,m} = \mathrm{E}_{a_m | \Omega_j}[-2 \log L(\Omega_j, a_m)]$ and $D_{j,m} = -2 \log L(\Omega_j, \tilde{a}_{j,m})$. So we can select the working model with the smallest value of the DIC. The CRM using the DIC is called as the CRM-DIC. The computations about the PPL and the DIC were performed using WinBUGS [19] in our example and simulation.

3.4. The dose-finding algorithm of the model-adaptive CRM

Given either of these above criteria, the dose-finding algorithm of the model-adaptive CRM proceeds as follows:

Step 1: Determine the starting dose. For example, for safety reasons the lowest dose level may be often chosen.
Step 2: Apply the CRM as shown in Section 2 to each of all elicited or possible working models at every observation following the inclusion of each patient or cohort of patients.
Step 3: Select one working model out of the working models that were updated by the CRM in Step. 2, through use of either of the PMP, the PPL or the DIC.
Step 4: Allocate the dose found by the selected working model in Step 3, to the next cohort of patients. Repeat Steps 2–4 until the last patient's data are observed.

We could add a stopping rule after Step 3 in the above algorithm, if necessary, in the same way as the original CRM. That is, the trial can be terminated for safety if given the first j included patients' data Ω_j, the lowest dose d_1 is too toxic:

$$\Pr(\psi(d_1, \tilde{a}_{j,m^*}) > \theta^*) > 0.9, \tag{5}$$

where m^* denotes the index of the selected working model. If the BMA approach is used, a stopping rule is given by

$$\sum_{m=1}^{M} \Pr(\psi(d_1, \tilde{a}_{j,m}) > \theta^*) \Pr(\mathsf{M}_m | \Omega_j) > 0.9. \tag{6}$$

4. A simulation study

We evaluated the performance of different model-selecting CRMs, i.e. the CRM based on the PMP (CRM-PMP), the CRM based on the PPL (CRM-PPL), and the CRM based on the DIC (CRM-DIC), through a simulation study with nine scenarios representing the underlying dose–toxicity relationships used in Section 3 of Yin and Yuan [15]. Two additional scenarios were studied, in which either a flat dose–toxicity curve up to the targeted MTD or a steep dose–toxicity curve were considered.

Under each scenario eight dose levels were chosen, and four sets of initial working models shown in the following:

$$\left\{ \begin{array}{ll} (0.02, 0.06, 0.08, 0.12, 0.20, 0.30, 0.40, 0.50), & \text{working model 1,} \\ (0.01, 0.05, 0.09, 0.14, 0.18, 0.22, 0.26, 0.30), & \text{working model 2,} \\ (0.10, 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80), & \text{working model 3,} \\ (0.20, 0.30, 0.40, 0.50, 0.60, 0.65, 0.70, 0.75), & \text{working model 4} \end{array} \right\}.$$

We refer to the individual CRMs using each of these toxicity probabilities, respectively, as CRM 1, CRM 2, CRM 3 and CRM 4. We set not only these CRMs but also the CRM based on the Bayesian model averaging (CRM-BMA), proposed by Yin and Yuan [15], as controls for comparisons with our proposed model-selecting CRMs. In the CRM-BMA, the prior working model probability for each of CRM 1–4 was assigned $\frac{1}{4}$, which was the same as that of Yin and Yuan [15]. The number of simulated trials was 1000. In each trial the first patient was allocated to d_1, and the maximum sample size with a stopping rule was 30. For the model-selecting CRMs and the CRM-BMA, the stopping rules were based on equations (5) and (6), respectively. For CRMs 1–4, equation (5) was used for each pre-specified working model. The performance was evaluated through the percentage of correct selection (PCS), i.e. the percentage of the recommended dose levels at the end of the trial, associated with the true MTD. In addition we estimated the mean-squared error (MSE) of the estimated toxicity probability to the true at the estimated MTD as

$$\sum_{N=1}^{1000} \frac{\{(\text{The estimated toxicity probability})_N - (\text{The true toxicity probability})\}^2}{1000}.$$

The target toxicity probability was $\theta^* = 0.3$. In Tables I and II, the scenarios, the PCS, the average number of DLTs, and the average sample size are given.

Scenario 1 had the MTD at d_7. The CRM-DIC as well as CRM 3 had a PCS over 70 per cent, and all the other designs had PCSs >53 per cent. In Scenario 2, CRM 3 behaved the best, whereas CRM 2 the least well. Our proposed model-selecting CRMs and the CRM-BMA lie between the two designs. In particular the CRM-DIC and the CRM-BMA showed PCSs about 45.6 per cent, with almost 30 per cent of patients treated at the MTD (shown in the supplementary

Table I. The percentage of correct selection (PCS), the average number of DLTs and the average sample size under Scenarios 1–6. The scenario and the PCS are given in italic and in bold, respectively.

	Percentage of correct selection										
Design	1	2	3	4	5	6	7	8	None	Ave. # tox.	Ave. # pats
Scenario 1	*0.02*	*0.03*	*0.04*	*0.06*	*0.08*	*0.1*	**0.3**	*0.5*			
CRM 1	0.0	0.0	0.0	0.0	0.4	14.3	**67.6**	17.7	0	8.6	30
CRM 2	0.0	0.0	0.0	0.0	2.0	20.7	**53.5**	23.8	0	8.8	30
CRM 3	0.0	0.0	0.0	0.3	1.6	18.7	**73.4**	6.0	0	6.1	30
CRM 4	0.0	0.0	0.0	0.2	1.6	19.6	**64.3**	14.3	0	6.4	30
CRM-PMP	0.0	0.0	0.0	0.3	1.0	15.9	**65.7**	17.1	0	8.4	30
CRM-PPL	0.0	0.0	0.0	0.2	1.5	18.1	**67.9**	12.3	0	6.2	30
CRM-DIC	0.0	0.0	0.0	0.2	0.6	20.4	**71.1**	7.7	0	7.4	30
CRM-BMA	0.0	0.0	0.0	0.1	1.0	17.0	**66.0**	15.9	0	8.0	30
Scenario 2	*0.02*	*0.06*	*0.08*	*0.12*	*0.2*	***0.3***	*0.4*	*0.5*			
CRM 1	0.0	0.0	0.1	1.4	23.0	**45.2**	25.1	5.2	0	9.3	30
CRM 2	0.0	0.0	0.5	6.1	24.5	**36.8**	23.1	9.0	0	9.2	30
CRM 3	0.0	0.0	0.5	4.1	30.4	**50.6**	14.0	0.4	0	7.3	30
CRM 4	0.0	0.0	0.5	5.0	34.1	**40.7**	17.3	2.4	0	7.0	30
CRM-PMP	0.0	0.1	0.6	4.3	20.5	**42.8**	26.7	5.0	0	9.0	30
CRM-PPL	0.0	0.0	0.5	4.7	31.6	**43.2**	18.2	1.8	0	7.0	30
CRM-DIC	0.0	0.0	0.2	5.2	26.6	**45.6**	19.3	3.1	0	8.4	30
CRM-BMA	0.0	0.0	0.2	3.2	24.8	**45.6**	22.3	3.9	0	8.5	30
Scenario 3	*0.06*	*0.15*	***0.3***	*0.55*	*0.6*	*0.65*	*0.68*	*0.7*			
CRM 1	0.8	26.7	**51.6**	18.8	1.5	0.2	0.0	0.0	0.4	10.8	29.9
CRM 2	0.1	24.1	**62.4**	12.4	0.4	0.1	0.0	0.0	0.5	10.2	29.9
CRM 3	0.3	19.2	**68.2**	11.4	0.4	0.1	0.0	0.0	0.4	9.4	29.9
CRM 4	0.3	19.2	**69.4**	10.2	0.4	0.0	0.0	0.0	0.5	9.0	29.9
CRM-PMP	0.4	20.7	**63.8**	13.5	1.0	0.2	0.0	0.0	0.4	10.2	29.9
CRM-PPL	0.4	19.9	**69.3**	10.0	0.0	0.0	0.0	0.0	0.4	9.0	29.9
CRM-DIC	0.1	21.2	**65.7**	12.1	0.4	0.1	0.0	0.0	0.4	10.1	29.9
CRM-BMA	0.5	22.1	**64.2**	12.2	0.5	0.0	0.1	0.0	0.4	10.0	29.9
Scenario 4	*0.2*	***0.3***	*0.4*	*0.5*	*0.6*	*0.65*	*0.7*	*0.75*			
CRM 1	20.3	**40.9**	21.4	8.6	1.2	0.1	0.0	0.0	7.5	10.4	28.2
CRM 2	13.4	**49.1**	23.5	4.1	0.7	0.0	0.0	0.0	9.2	9.8	28.1
CRM 3	15.5	**45.7**	26.5	5.1	0.3	0.0	0.0	0.0	6.9	9.5	28.3
CRM 4	16.8	**45.7**	23.5	4.5	0.3	0.1	0.0	0.0	9.1	8.9	27.8
CRM-PMP	15.4	**45.1**	25.2	4.8	0.0	0.0	0.0	0.0	9.5	10.0	27.6
CRM-PPL	16.7	**46.8**	23.7	4.6	0.5	0.0	0.0	0.0	7.7	9.4	28.3
CRM-DIC	16.0	**45.1**	25.3	4.3	0.7	0.1	0.0	0.0	8.5	10.1	28.1
CRM-BMA	18.3	**45.3**	23.4	4.3	0.7	0.0	0.0	0.0	8.0	10.1	28.1
Scenario 5	*0.1*	*0.2*	***0.3***	*0.4*	*0.5*	*0.6*	*0.7*	*0.8*			
CRM 1	1.3	27.7	**31.8**	30.5	7.1	0.4	0.1	0.0	1.1	10.6	29.7
CRM 2	1.2	27.1	**41.4**	23.1	4.9	0.8	0.0	0.0	1.5	9.9	29.6
CRM 3	1.0	23.0	**45.9**	25.0	3.6	0.3	0.0	0.0	1.2	9.4	29.7
CRM 4	1.2	22.4	**48.5**	24.0	2.2	0.1	0.0	0.0	1.6	8.9	29.6
CRM-PMP	1.2	23.7	**46.3**	23.0	4.0	0.5	0.0	0.0	1.3	10.0	29.7
CRM-PPL	1.3	24.6	**46.7**	21.8	3.8	0.2	0.0	0.0	1.6	8.9	29.6
CRM-DIC	1.4	22.1	**46.8**	24.6	3.8	0.0	0.0	0.0	1.3	9.8	29.7
CRM-BMA	0.9	23.5	**43.8**	24.3	6.0	0.3	0.0	0.0	1.2	9.8	29.7
Scenario 6	*0.02*	*0.03*	*0.05*	*0.07*	***0.3***	*0.5*	*0.7*	*0.8*			
CRM 1	0.0	0.0	0.0	11.3	**70.6**	17.8	0.3	0.0	0	9.7	30
CRM 2	0.0	0.0	0.1	20.5	**57.1**	20.5	1.7	0.1	0	9.4	30
CRM 3	0.0	0.0	0.0	11.1	**73.2**	15.7	0.0	0.0	0	8.4	30
CRM 4	0.0	0.0	0.0	11.5	**71.1**	16.9	0.5	0.0	0	7.9	30
CRM-PMP	0.0	0.0	0.0	11.8	**69.7**	18.2	0.3	0.0	0	9.4	30
CRM-PPL	0.0	0.0	0.0	11.9	**70.6**	17.0	0.5	0.0	0	7.9	30
CRM-DIC	0.0	0.0	0.0	12.8	**70.6**	16.5	0.0	0.1	0	8.9	30
CRM-BMA	0.0	0.0	0.0	13.7	**68.8**	17.4	0.1	0.0	0	9.0	30

Table II. The percentage of correct selection (PCS), the average number of DLTs and the average sample size under Scenarios 1 to 6. The scenario and the PCS are given in italic and in bold, respectively.

	Percentage of correct selection										
Design	1	2	3	4	5	6	7	8	None	Ave. # tox.	Ave. # pats
Scenario 7	*0.03*	*0.07*	*0.1*	*0.15*	*0.2*	***0.3***	*0.5*	*0.7*			
CRM 1	0.0	0.0	0.2	3.7	28.4	**54.9**	12.6	0.2	0	9.2	30
CRM 2	0.0	0.0	1.0	9.9	30.8	**41.5**	15.2	1.6	0	9.1	30
CRM 3	0.0	0.0	1.0	6.7	30.5	**54.0**	7.8	0.0	0	7.3	30
CRM 4	0.0	0.0	0.9	9.6	35.5	**42.6**	11.3	0.1	0	7.1	30
CRM-PMP	0.0	0.0	0.5	6.9	27.8	**52.5**	12.0	0.3	0	8.8	30
CRM-PPL	0.0	0.0	1.1	8.9	32.8	**45.2**	11.8	0.2	0	7.1	30
CRM-DIC	0.0	0.0	0.7	8.3	32.3	**50.6**	8.1	0.0	0	8.1	30
CRM-BMA	0.0	0.0	0.4	6.1	27.9	**53.3**	12.2	0.1	0	8.5	30
Scenario 8	*0.02*	*0.03*	*0.05*	*0.06*	*0.07*	*0.09*	*0.1*	***0.3***			
CRM 1	0.0	0.0	0.0	0.0	0.0	0.8	13.3	**85.9**	0	7.1	30
CRM 2	0.0	0.0	0.0	0.0	0.1	1.3	12.6	**86.0**	0	7.3	30
CRM 3	0.0	0.0	0.0	0.3	1.5	4.9	31.3	**62.0**	0	4.4	30
CRM 4	0.0	0.0	0.0	0.3	1.2	3.2	24.4	**70.9**	0	5.0	30
CRM-PMP	0.0	0.0	0.0	0.1	0.7	1.6	12.7	**84.9**	0	7.0	30
CRM-PPL	0.0	0.0	0.0	0.3	1.1	3.2	25.9	**69.5**	0	4.7	30
CRM-DIC	0.0	0.0	0.0	0.2	0.5	2.7	27.1	**69.5**	0	6.3	30
CRM-BMA	0.0	0.0	0.0	0.1	0.7	1.1	14.8	**83.3**	0	6.7	30
Scenario 9	*0.4*	*0.5*	*0.6*	*0.7*	*0.8*	*0.9*	*0.95*	*0.99*			
CRM 1	35.7	4.7	0.2	0.1	0.0	0.0	0.0	0.0	59.3	8.6	18.1
CRM 2	19.5	6.5	0.2	0.0	0.0	0.0	0.0	0.0	73.8	8.0	16.5
CRM 3	37.2	7.2	0.4	0.0	0.0	0.0	0.0	0.0	55.2	8.5	18.7
CRM 4	30.4	6.0	0.2	0.0	0.0	0.0	0.0	0.0	63.4	7.6	16.9
CRM-PMP	29.1	8.0	0.4	0.0	0.0	0.0	0.0	0.0	62.5	8.3	17.0
CRM-PPL	31.2	5.2	0.2	0.0	0.0	0.0	0.0	0.0	63.4	8.2	18.2
CRM-DIC	28.4	8.1	0.6	0.0	0.0	0.0	0.0	0.0	62.9	8.4	18.1
CRM-BMA	32.2	6.3	0.0	0.1	0.0	0.0	0.0	0.0	61.4	8.4	18.3
Scenario 10	*0.2*	*0.21*	*0.22*	*0.23*	*0.24*	*0.25*	***0.3***	*0.35*			
CRM 1	1.8	2.8	2.7	5.1	12.4	16.8	**25.5**	27.9	5.0	8.1	28.7
CRM 2	1.9	4.5	5.4	9.0	10.9	11.1	**18.4**	33.5	5.3	8.2	28.7
CRM 3	2.8	6.4	8.1	14.1	16.3	24.7	**18.2**	4.5	4.9	6.9	28.7
CRM 4	2.3	7.9	12.5	19.1	16.3	14.4	**13.9**	6.7	6.9	6.7	28.3
CRM-PMP	1.9	4.7	5.7	8.0	9.2	14.5	**22.1**	28.1	5.8	7.9	28.6
CRM-PPL	2.6	8.5	11.4	19.1	16.3	16.3	**13.4**	6.2	6.2	6.8	28.5
CRM-DIC	1.7	5.2	6.1	9.9	10.7	18.4	**22.1**	20.8	5.1	7.8	28.7
CRM-BMA	5.9	4.3	6.4	9.5	12.2	17.9	**20.0**	22.6	1.2	7.6	28.8
Scenario 11	*0.01*	*0.05*	*0.10*	***0.3***	*0.5*	*0.6*	*0.7*	*0.8*			
CRM 1	0.0	0.7	15.1	**63.6**	20.1	0.4	0.1	0.0	0.0	10.4	30
CRM 2	0.0	0.3	18.2	**59.0**	20.0	2.3	0.2	0.0	0.0	9.7	30
CRM 3	0.0	0.0	13.6	**69.8**	16.3	0.3	0.0	0.0	0.0	9.0	30
CRM 4	0.0	0.0	13.1	**70.9**	15.8	0.2	0.0	0.0	0.0	8.5	30
CRM-PMP	0.0	0.0	14.9	**66.6**	17.3	1.2	0.0	0.0	0.0	9.8	30
CRM-PPL	0.0	0.2	14.1	**67.8**	17.5	0.4	0.0	0.0	0.0	8.5	30
CRM-DIC	0.0	0.0	13.9	**70.2**	15.1	0.8	0.0	0.0	0.0	9.5	30
CRM-BMA	0.0	0.1	14.3	**65.1**	20.1	0.4	0.0	0.0	0.0	9.4	30

material).[‡] In Scenario 3, all designs except CRM 1 had high PCSs (>62 per cent). In Scenarios 4 and 5, all designs except CRM 1 had close PCSs. If we had to compare among the model-selecting CRMs and the CRM-BMA, the CRM-PPL behaved the best overall. CRM 2 performed the least well in Scenario 6, associated with a PCS almost 10 per cent lower than the other designs. In Scenario 7, the CRM-BMA had a relatively higher PCS, compared with the model-selecting CRMs. In Scenario 8, the CRM-PMP and the CRM-BMA had a relatively higher PCS, but the CRM-PPL and the CRM-DIC gave lower PCSs compared with other designs. In Scenario 9, in which d_1 was associated with a toxicity probability higher than the target, showed that the used stopping rules could stop the trial earlier with

[‡]*Supporting information may be found in the online version of this article.*

no dose recommendation in over 55 per cent of trials. Using a flat dose–toxicity relationship as in Scenario 10 showed that all designs except CRM-DIC overestimated or underestimated the MTD. In Scenario 11, in which the underlying dose–toxicity curve had a steep slope, the CRM-DIC as well as CRM 4 yielded a PCS >70 per cent. For all the scenarios, the average number of observed toxicities was similar within each scenario.

The more spaced out the dose levels, in terms of coding, the easier for the method to distinguish one level from another. In practice our suggestion would be to roughly divide the interval (0,1) into $I+1$ equal intervals (for the examples here taken from Yin and Yuan [15] this is $8+1=9$). If we assess the performance in terms of the percentage of recommendations, at the MTD, one level below the MTD and one level above the MTD, then there is very little to choose between the different models. They behave much the same. If we only consider the percentage recommendation at the MTD itself, then the working model CRM 2, in scenarios 1, 2 and 6, while performing well, does perform less well than other model choices. For the model CRM 2, the spacing is much tighter than for the other choices and this is most likely the explanation for the differences in performance. Model choice CRM 1 performs less well than the other choices in scenarios 3, 4 and 5, and, again, in these cases, the MTD was located in a region where the model choice bunched up the dose levels. At other scenarios, when the MTD was located in an area corresponding to more spread determined by the model, the performance was comparable. For model choice CRM 3 and CRM 4, where the levels were spread widely, the performance was in all cases satisfactory. The model-selecting CRMs and the CRM-BMA may help reduce the influence of a less than fortunate model specification. In practice, when we utilize the model-selecting CRMs or the CRM-BMA, one way to identify a less than optimal working model is to plot values of the used criterion, say, PMP, PPL, or DIC, for each pre-specified working model, against accumulating patients or cohorts (see [15]). A great

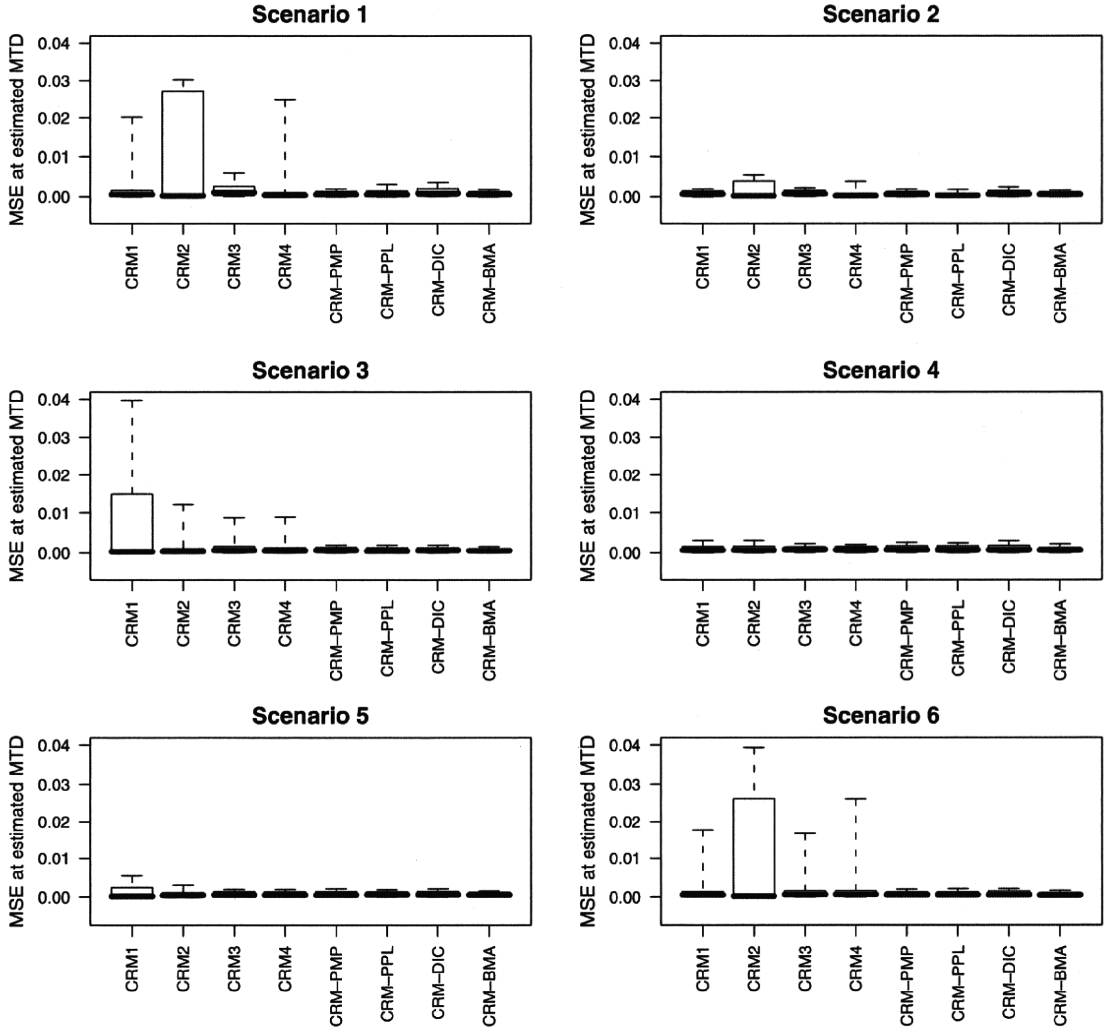


Figure 1. Comparison of the MSEs at the estimated MTD under Scenarios 1–6.

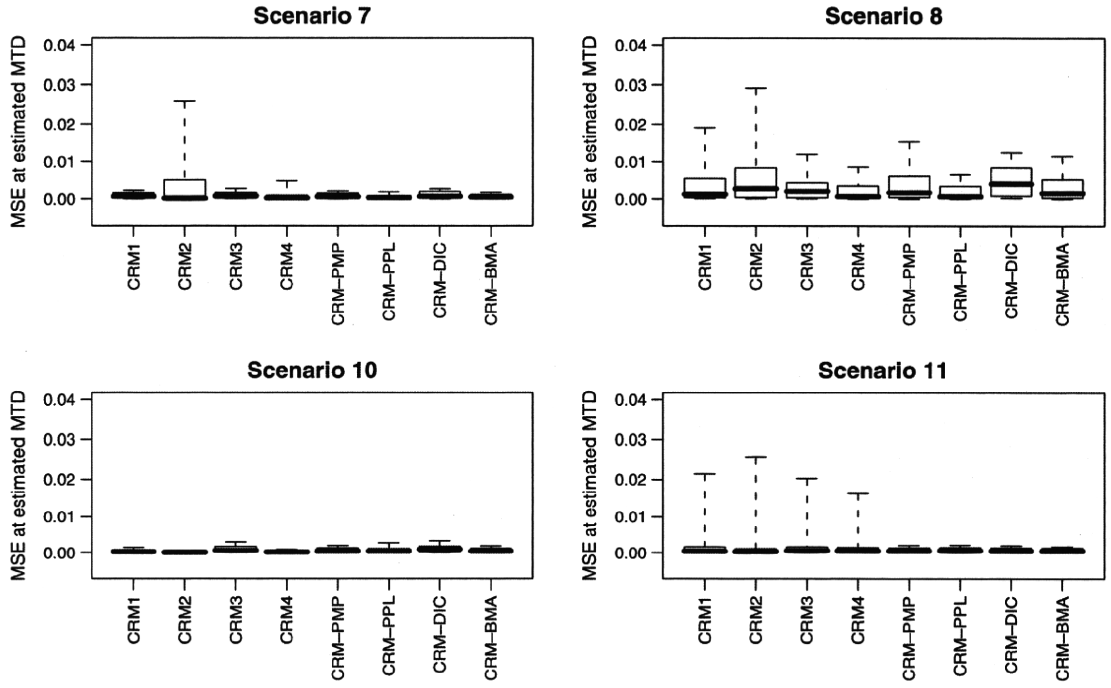


Figure 2. Comparison of the MSEs at the estimated MTD under Scenarios 7–11 (except Scenario 9).

difference in the values exceeding some threshold might give us a reason for being cautious in the use of a particular working model during the course of a trial.

Figures 1 and 2 show the boxplots of the MSEs at the estimated MTD under each scenario (except Scenario 9 with early stopping). Our proposed model-selecting CRMs and the CRM-BMA tended to yield the narrower quartile ranges than the worst-performed CRM.

5. Conclusions and discussions

The results of the above section show that the poor pre-specification of the toxicity probabilities in a single CRM design could have a negative impact on both the MTD estimation and the dose allocation to each patient. The model-selecting CRMs as well as the CRM-BMA may help to reduce the potential for errors following such specification.

We have investigated a model-selecting CRM design, in which we use multiple working models together with model selection, based on the Bayesian model selection criteria, such as the PMP, the PPL and the DIC. The results of the simulation study show that all of the three proposed designs can provide a compromise solution between the best-performing and worst-performing model. In other words, it may be possible to improve the robustness of the CRM to poor pre-specification of a single working model. The Bayesian model averaging CRM will average several working models into a single one unlike the model-selecting CRMs. From the results of the simulation study, it appears that the former performs better under some scenarios and the latter under others. Roughly, over the class of working models, there are two kinds of simple summaries, the mean and the mode; the Bayesian model averaging CRM being the mean and the model-selecting CRMs being the mode. An immediate question is which one of these summaries is to be preferred. For the cases we have studied here there is no clear cut answer to this and no clear cut winner. Although only speculative, we believe that if it were possible to identify and include all possible dose–toxicity curves, as well as all potential working models, in a comparative study, there would still be no clear cut winner. The comparison is more complex but not fundamentally all that different from a comparison of summary measures for some given unimodal random variable U with density $f(u)$. We could consider E(U) or $u_m : f(u_m) > f(u), u \in \mathcal{U}$ where $\mathcal{U}$ is the support of U. If the variable is symmetric then the two measures coincide. If positively skewed then $E(U)$ will be more sensitive to larger values of u than u_m. This may be an argument in its favor when we would like the impact of the larger values to be given due weight but an argument against it when these larger values are considered potentially too influential. In that case u_m would be much less sensitive to the larger values. At the time of writing not enough is known about the distribution of the variable of interest—in this case the set of potential working models—to be able to provide firm recommendations. As more becomes known through real studies and accumulated experience, it may be possible in the future to provide sharper indications.

Although it appears too difficult to rank the three proposed model-selecting CRMs, our simulations provide some insight into their different characteristics. For example, when the true dose–toxicity relationship is quite flat up to the targeted MTD (in this situation some existing methods fail in the MTD estimation [20]), the model-selecting CRM based on the DIC works well. The model-selecting CRM based on the PPL appears to provide a conservative dose escalation, resulting in the number of patients allocated to the higher more toxic dose levels being relatively smaller than with the other model-selecting techniques. This may be a result of the PPL being based on the posterior predictive distribution allowing for larger variability than the posterior distribution. The model-selecting CRM based on the PMP could be quite similar in performance to the model averaging CRM because both use the PMP.

An interpretation of the working model in terms of Bayesian prior point estimates of the probabilities of toxicity is quite common in this area. Although such an interpretation is not only needed but may not correspond in practice to anything very concrete (since operating characteristics and final recommendations remain invariant to these prior point estimates being raised to any positive power), many clinicians and experimenters are reassured by the interpretation. When this is so, an additional reassurance can be obtained by using several working models as opposed to a single one, in particular in cases when there are differences in opinion among investigators. Rather than force the choice to any one particular working model it is possible to say that all the suggested potential working models can be employed and then, as the observations are accrued, the method itself will help select the most plausible given this information. In reality, this will still end up being equivalent to having worked with a particular simple model choice. However, for those investigators who may be concerned at the apparent arbitrariness of working model selection, and the worry that this may impact the relative performance of the method, the use of several models enables them to gain confidence in the resulting MTD estimate, based as it is on a number of rival models.

Acknowledgements

This research was supported partially by the Health Labour Sciences Research Grant (Research on Measures for Intractable Diseases) of the Ministry of Health, Labour and Welfare, and by Grant-in-Aid for Researchers, Hyogo College of Medicine. We thank the two referees and the editors for their helpful comments that improved the quality of the paper.

References

1. Storer BE. Design and analysis of phase I clinical trials. *Biometrics* 1989; **45**:925–937.
2. Lin Y, Shih WJ. Statistical properties of the traditional algorithm-based designs for phase I cancer clinical trials. *Biostatistics* 2001; **2**:203–215.
3. Ivanova A, Montazer-Haghighi A, Mohanty SG, Durham SD. Improved up-and-designs for Phase I trials. *Statistics in Medicine* 2003; **22**:69–82.
4. Ivanova A. Escalation up-and-down and $A+B$ designs for dose-finding trials. *Statistics in Medicine* 2003; **25**:3668–3678.
5. O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase 1 clinical trials in cancer. *Biometrics* 1990; **46**:33–48.
6. Whitehead J, Williamson D. Bayesian decision procedures based on logistic regression models for dose-finding studies. *Journal of Biopharmaceutical Statistics* 1998; **8**:445–467.
7. Babb J, Rogatko A, Zacks S. Cancer phase I clinical trials: efficient dose escalation with overdose control. *Statistics in Medicine* 1998; **17**:1103–1120.
8. Shen LZ, O'Quigley J. Consistency of continual reassessment method under model misspecification. *Biometrika* 1996; **83**:395–405.
9. O'Quigley J. Another look at two phase I clinical trial designs. *Statistics in Medicine* 1999; **18**:2683–2690.
10. Iasonos A, Wilton AS, Riedel ER, Seshan VE, Spriggs DR. A comprehensive comparison of the continual reassessment method to the standard 3+3 dose escalation scheme in Phase I dose-finding studies. *Clinical Trials* 2008; **5**:465–477.
11. O'Quigley J, Zohar S. Retrospective robustness of the continual reassessment method. *Journal of Biopharmaceutical Statistics* 2010; **20**(5):1013–1025.
12. O'Quigley J, Shen LZ. Continual reassessment method: a likelihood approach. *Biometrics* 1996; **52**:673–684.
13. Cheung YK, Chappell R. A simple technique to evaluate model sensitivity in the continual reassessment method. *Biometrics* 2002; **58**:671–674.
14. Lee SM, Cheung YK. Model calibration in the continual reassessment method. *Clinical Trials* 2009; **6**:227–238.
15. Yin G, Yuan Y. Bayesian model averaging continual reassessment method in phase I clinical trials. *Journal of the American Statistical Association* 2009; **104**:954–968.
16. Gelfand A, Ghosh S. Model choice: a minimum posterior predictive loss approach. *Biometrika* 1998; **85**:1–11.
17. Spiegelhalter D, Best N, Carlin B, van der Linde A. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* 2002; **64**:583–639.
18. Kass RE, Raftery AE. Bayes factors and model uncertainty. *Journal of the American Statistical Association* 1995; **90**:773–795.
19. Lunn DJ, Thomas A, Best NG, Spiegelhalter DJ. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 2000; **10**:325–337.
20. Cheung YK. Sequential implementation of stepwise procedures for identifying the maximum tolerated dose. *Journal of the American Statistical Association* 2007; **102**:1448–1461.

Pulmonary hypertension predicts adverse cardiac events after restrictive mitral annuloplasty for severe functional mitral regurgitation

Satoshi Kainuma, MD,[a] Kazuhiro Taniguchi, MD, PhD,[a] Koichi Toda, MD, PhD,[a] Toshihiro Funatsu, MD, PhD,[a] Haruhiko Kondoh, MD, PhD,[a] Masami Nishino, MD, PhD,[b] Takashi Daimon, PhD,[c] and Yoshiki Sawa, MD, PhD[d]

Objectives: Pulmonary hypertension (PH) is an indicator of a poor prognosis in patients with dilated cardiomyopathy. Few studies have investigated the prognostic role of PH in patients undergoing restrictive mitral annuloplasty (RMA) for severe functional mitral regurgitation secondary to advanced cardiomyopathy.

Methods: A total of 46 patients undergoing RMA were classified into 3 groups on the basis of the Doppler-derived systolic pulmonary artery pressure (PAP) at baseline. Of the 46 patients, 19 had a systolic PAP less than 40 mm Hg (mild PH group), 17 had a systolic PAP of 40 to 60 mm Hg (moderate PH group), and 10 had a systolic PAP greater than 60 mm Hg (severe PH group).

Results: Postoperative cardiac catheterization showed that the RMA procedure resulted in a significant reduction of the left ventricular (LV) preload and improvements in LV systolic function in all 3 groups, along with the relief of symptoms. During the follow-up period (mean 36 ± 19 months), cardiac death occurred in 6 patients, readmission because of heart failure in 3, and fatal arrhythmia in 1. The rate of freedom from these cardiac events at 3 years was 93% ± 7%, 88% ± 8%, and 56% ± 17% in the mild, moderate, and severe PH groups ($P < .001$). Serial echocardiography showed that significant LV reverse remodeling occurred in 89%, 71%, and 25% of the mild, moderate, and severe PH groups, respectively. Multivariate Cox regression analysis identified severe PH (systolic PAP > 60 mm Hg) as a significant predictor of adverse cardiac events, as well as LV remodeling after RMA.

Conclusions: Noninvasive assessment of preoperative PH has a prognostic value in patients undergoing RMA for severe functional mitral regurgitation secondary to advanced cardiomyopathy. (J Thorac Cardiovasc Surg 2010;■:1-10)

It is well known that pulmonary hypertension (PH) is an indicator of a poor prognosis in patients with dilated cardiomyopathy and functional mitral regurgitation (MR),[1] as well as in heart transplant recipients.[2] However, whether PH has the same prognostic value in patients who have undergone surgery for functional MR complicated by advanced cardiomyopathy is unknown.

Since Bolling and colleagues[3] first reported the feasibility of surgery for uncontrollable severe MR in patients with end-stage cardiomyopathy, restrictive mitral annuloplasty (RMA) has become the preferred surgical treatment of this condition. Several factors have been shown to be predictors of poor outcome after RMA.[4-8] However, few investigations have assessed the influence of PH on symptomatic improvements, left ventricular (LV) function, or survival after RMA in patients with advanced cardiomyopathy.

In the present study, we investigated whether PH has a prognostic value in patients undergoing RMA for functional MR. In addition, we determined which parameters might function as predictors of postoperative outcome and reverse LV remodeling in those patients, with a focus on preoperative PH.

MATERIALS AND METHODS

Patients

We examined the records of 65 consecutive patients who had undergone RMA for functional MR at our institution from March 2004 to December 2009. Of those, 46 patients were chosen as study subjects, according to the following inclusion criteria: 1) chronic heart failure with New York Heart Association (NYHA) functional class III or IV and a history of at least 1 hospitalization; 2) advanced LV remodeling, defined as a LV ejection

From the Departments of Cardiovascular Surgery[a] and Cardiology,[b] Japan Labor Health and Welfare Organization, Osaka Rosai Hospital, Osaka, Japan; Department of Biostatistics,[c] Hyogo College of Medicine, Hyogo, Japan; Department of Cardiovascular Surgery,[d] Osaka University Graduate School of Medicine, Osaka, Japan.

This research was supported by research funds to promote the hospital function of the Japan Labor Health and Welfare Organization.

Disclosures: Authors have nothing to disclose with regard to commercial support.

This study was presented in part at the American Heart Association Scientific Sessions 2009, November 14–18, 2009, Orlando, Florida.

Received for publication Aug 27, 2010; revisions received Nov 5, 2010; accepted for publication Nov 19, 2010.

Address for reprints: Kazuhiro Taniguchi, MD, Department of Cardiovascular Surgery, Japan Labor Health and Welfare Organization, Osaka Rosai Hospital, 1179-3 Nagasone-cho, Kita-ku, Sakai, Osaka 591-8025 Japan (E-mail: kataniguchi-cvs@orh.go.jp).

0022-5223/$36.00

doi:10.1016/j.jtcvs.2010.11.031

Abbreviations and Acronyms

BNP = brain natriuretic peptide
LV = left ventricular
LVEDD = left ventricular end-diastolic dimension
LVEF = left ventricular ejection fraction
MR = mitral regurgitation
NYHA = New York Heart Association
PAP = pulmonary artery pressure
PH = pulmonary hypertension
PVR = pulmonary vascular resistance
RMA = restrictive mitral annuloplasty
TR = tricuspid regurgitation

fraction (LVEF) of less than 40% and LV end-systolic volume index greater than 60 ml/m^2, as shown by left ventriculography; and 3) severe MR caused by restrictive leaflet motion secondary to global LV dilatation. Patients with a lesser degree of LV remodeling and ischemic MR secondary to a regional LV deformity due to inferior/posterior myocardial infarction were excluded from the present study. The patients with recent myocardial infarction (< 3 months), organic MR, rheumatic mitral disease, or a known noncardiac cause of PH and those who had undergone concomitant surgical ventricular reconstruction were not included in the present study.

To determine whether the degree of preoperative PH was associated with the postoperative outcome, the patients were classified into 3 groups according to the Doppler-derived systolic pulmonary artery pressure (PAP) at baseline. Of the 46 patients, 19 had a systolic PAP of less than 40 mm Hg (mild PH group), 17 had a systolic PAP of 40 to 60 mm Hg (moderate PH group), and 10 had a systolic PAP greater than 60 mm Hg (severe PH group).

The clinical characteristics and surgical data are listed in Table 1. No significant differences were found in gender, body surface area, NYHA functional class, or prevalence of complications among the 3 groups. The patients in the severe PH group had had a longer duration of heart failure than those in the other groups, although they were younger than the patients in the moderate PH group.

The related institutional ethics committees approved the present study and waived the need for individual consent for the retrospective analysis. Each patient provided written informed consent for the procedure before surgery.

Echocardiographic Measurements and Calculations

Two-dimensional and Doppler transthoracic echocardiography were performed at baseline, at 1 month after surgery (mean 28 ± 5 days), and annually thereafter. The preoperative (baseline) and postoperative echocardiographic examinations at 1 month after surgery were performed within 1 day of cardiac catheterization. Transesophageal echocardiography was also performed within 1 week before surgery to confirm the severity and precise mechanism of MR in all patients. All echocardiographic studies were performed using commercially available 3.75-MHz transducers (Toshiba, Tokyo, Japan, and Hewlett-Packard Sonos) by the same echocardiographic expert examiner (S.F.), who was unaware of the clinical status of the patients.

LV Function and Left Atrial Dimensions

The LV end-diastolic dimension (LVEDD), LV end-systolic dimension, and left atrial dimension was determined from 2-dimensional echocardiographic images in the parasternal long-axis views. The LVEF was calculated using Simpson's method with 2 apical views.

Doppler-Derived Systolic PAP

The systolic PAP was calculated by adding the systolic pressure gradient across the tricuspid valve derived from the tricuspid regurgitation (TR) to the estimated right atrial pressure.[9,10] The tricuspid regurgitant signal was recorded by continuous-wave Doppler echocardiography, and its maximal velocity was measured. Using the simplified Bernoulli equation, the pressure gradient across the tricuspid valve was calculated. The right atrial pressure was estimated using the dimension of the inferior vena cava and the response to changes in respiration.[11] In the present patients, a tricuspid regurgitant signal was detected in all patients during the baseline examination and in the great majority of patients during the follow-up examinations.

Mitral and Tricuspid Valve Measurements

The severity of MR and TR was graded semiquantitatively from the color flow Doppler data. In our routine assessment, MR severity was characterized as none (0), trivial (1+), mild (2+), moderate (3+), or severe (4+), depending on how far beyond the mitral valve the regurgitation extended into the left atrium. The tenting height was measured between the line connecting the annular hinge points and the leaflet coaptation point, and the coaptation length was measured directly. The area of the mitral valve orifice was determined by direct planimetry or using the pressure half-time method, and the mean transmitral diastolic gradient was calculated using the Bernoulli equation determined from continuous-wave Doppler echocardiography.

Cardiac Catheterization and Hemodynamic Measurements

Cardiac catheterization was performed before and 1 month after surgery to measure the following routine cardiac hemodynamic parameters: LV end-diastolic volume index, LV end-systolic volume index, LV systolic pressure, LV end-diastolic pressure, pulmonary capillary wedge pressure, systolic and mean PAP, transmitral diastolic pressure gradient (pulmonary capillary wedge pressure minus the LV end-diastolic pressure), right atrial pressure, systemic vascular resistance ([mean arterial pressure minus right atrial pressure] multiplied by 80 and divided by the cardiac output), and pulmonary vascular resistance (PVR) ([mean PAP minus pulmonary capillary wedge pressure] multiplied by 80 and divided by the cardiac output). Our catheterization technique has been previously described.[12]

Surgical Procedures

A median sternotomy was performed under a mild hypothermic cardiopulmonary bypass, with intermittent cold blood cardioplegia. Mitral valve surgery was performed through a trans-septal superior approach. Carpentier-Edwards physio rings (Carpentier Ring, Edwards Lifesciences, Irvine, Calif) were used for all RMA procedures. The ring size was determined after careful measurement of the intercommissural distance and the height of the anterior leaflet and then downsizing by 2 to 3 sizes. No other adjunct procedures were performed on the valve itself. No significant differences were found in regard to the surgical procedures among the groups, including the size of the mitral annulus ring implanted and frequency of the concomitant procedures (Table 1).

Clinical Follow-up

Clinical follow-up examinations were completed for the 43 operative survivors, with a mean duration of 36 ± 19 months (range 5–77 months). After surgery, the patients were treated with standard heart failure medications, including angiotensin-converting enzyme inhibitors or angiotensin-II receptor blockers, β-blockers, and diuretics. Every 6 to 12 months, they were assessed in our department and by their primary cardiologist. The functional status was assessed according to the NYHA criteria for

TABLE 1. Clinical characteristics and surgical data

Parameter	All (n = 46)	Mild (n = 19)	Moderate (n = 17)	Severe (n = 10)	*P* Value
Age (y)	64 ± 8	62 ± 9	68 ± 6*	60 ± 7	.02
Men (%)	35 (76)	13 (68)	14 (82)	8 (80)	NS
Body surface area (m^2)	1.7 ± 0.2	1.6 ± 0.2	1.7 ± 0.2	1.6 ± 0.2	NS
NYHA functional class	3.2 ± 0.4	3.2 ± 0.4	3.1 ± 0.3	3.2 ± 0.4	NS
Duration of HF (mo)	25 ± 15	18 ± 9*	24 ± 14*	39 ± 15	< .01
Ischemic etiology (%)	32 (70)	14 (74)	12 (71)	6 (60)	NS
Hypertension	30 (65)	13 (68)	12 (71)	5 (50)	NS
Diabetes	17 (37)	8 (42)	6 (35)	3 (30)	NS
Hyperlipidemia	20 (43)	10 (0)	6 (35)	4 (40)	NS
COPD	7 (15)	1 (5)	3 (18)	3 (30)	NS
Chronic renal disease	16 (35)	7 (37)	3 (18)	6 (60)	NS
Peripheral vascular disease	6 (13)	0 (0)	3 (18)	3 (30)	NS
Cerebrovascular accident	14 (30)	4 (21)	6 (35)	4 (40)	NS
Atrial fibrillation	21 (46)	9 (47)	8 (47)	4 (40)	NS
Previous CABG	1 (2)	0 (0)	0 (0)	1 (10)	NS
Previous PCI	13 (28)	5 (26)	5 (29)	3 (30)	NS
β-Blockers	33 (72)	17 (89)	11 (65)	5 (50)	NS
ACE inhibitors	9 (20)	3 (16)	4 (24)	2 (20)	NS
ARB	17 (37)	9 (47)	7 (41)	1 (10)	NS
Nitrate	12 (26)	6 (32)	3 (18)	3 (30)	NS
Diuretics	34 (74)	11 (58)	15 (82)	8 (80)	NS
Surgical data					
CPB (min)	240 ± 71	222 ± 43	252 ± 69	252 ± 109	NS
ACC (min)	133 ± 43	130 ± 40	147 ± 42	114 ± 43	NS
Physio ring (mm)					
24	37 (80)	15 (79)	15 (82)	7 (70)	NS
26	9 (20)	4 (21)	2 (12)	3 (30)	NS
Concomitant procedure					
CABG	26 (57)	11 (58)	10 (59)	5 (50)	NS
TAP	43 (93)	17 (89)	16 (94)	10 (100)	NS
Modified maze	21 (46)	9 (47)	8 (47)	4 (40)	NS

NYHA, New York Heart Association; *HF,* heart failure; *COPD,* chronic obstructive pulmonary disease; *CABG,* coronary artery bypass grafting; *PCI,* percutaneous coronary intervention; *ACE,* angiotensin-converting enzyme; *ARB,* angiotensin-II receptor blocker; *CPB,* cardiopulmonary bypass; *ACC,* aortic crossclamp; *TAP,* tricuspid annuloplasty.
*$P < .05$ versus severe PH.

symptoms of heart failure and serum brain natriuretic peptide (BNP) level. A retrospective review of the medical records of these patients was performed for the preoperative and postoperative data, and the current information was obtained by interviewing the patient or the referring cardiologist. The postoperative adverse cardiac events included cardiac death, myocardial infarction, endocarditis, thromboembolism, reoperation for recurrent MR, readmission for heart failure, and fatal arrhythmia.

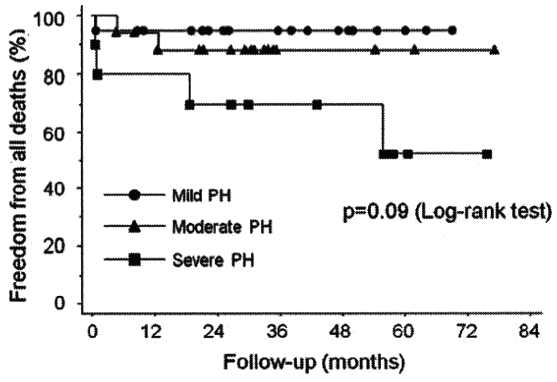
Statistical Analysis

The values for continuous variables are expressed as the mean ± standard deviation. The Kruskal-Wallis and chi-square tests were used to compare the pre- and postoperative values among the 3 groups. The functional and echocardiographic variables over time were compared using repeated measures analysis of variance, followed by a Bonferroni test for individual significant differences. Univariate analysis of the predictors of adverse cardiac events was performed using a Cox proportional hazard model (see Appendix). The factors with $P < .1$ were then entered appropriately into a multivariate model. The results are summarized as hazard ratios and 95% confidence intervals. Stepwise logistic regression analysis was performed to identify predictors for failure of reverse LV remodeling (see Appendix). Kaplan-Meier and log-rank analyses were performed to compare survival and the freedom from adverse cardiac events. Correlations between the Doppler-derived and catheter-measured systolic PAP were tested using linear correlation analysis. A Bland-Altman analysis was used to further determine the agreement between the 2 modalities by calculating the bias (mean difference) and 95% limits of agreement (± 2 standard deviation).[13] $P < .05$ was considered statistically significant. The statistical analyses were performed using JMP, version 7.0 (SAS Institute, Cary, NC).

RESULTS

Survival

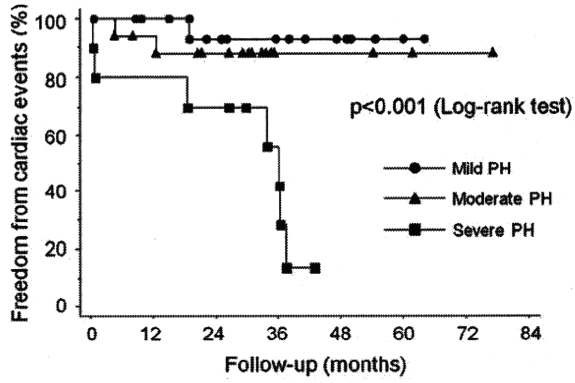
In the mild PH group, 1 early noncardiac death and 1 late fatal arrhythmia occurred. In the moderate PH group, 2 late cardiac deaths occurred. In the severe PH group, 2 early cardiac deaths, 2 late cardiac deaths, and 3 late readmissions because of heart failure occurred. None of the patients required reoperation for MR recurrence or endocarditis or presented with myocardial infarction or cerebrovascular or thromboembolic events during the follow-up period. The actuarial survival rate free from all deaths at 3 years was 95% ± 5%, 88% ± 8%, and 70% ± 15% in the mild, moderate, and severe PH groups, respectively ($P = .09$;



Patients at risk

	0	12	24	36	48	60	72 (month)
Mild PH	19	16	13	10	6	3	0
Moderate PH	17	15	12	3	3	2	1
Severe PH	10	8	7	5	4	2	1
Total	46	39	32	18	13	7	2

FIGURE 1. Actuarial survival rates. *Circles*, *triangles*, and *squares* indicate mild, moderate and severe groups, respectively. *PH*, Pulmonary hypertension. Numbers at bottom indicate patients at risk at each interval.



Patients at risk

	0	12	24	36	48	60	72 (month)
Mild PH	19	16	12	9	5	2	0
Moderate PH	17	15	12	3	3	2	1
Severe PH	10	8	7	4	0	0	0
Total	46	39	31	16	8	4	1

FIGURE 2. Freedom from adverse cardiac events. *Circles*, *triangles*, and *squares* indicate mild, moderate, and severe groups, respectively. *PH*, Pulmonary hypertension. Numbers at bottom indicate patients at risk at each interval.

Figure 1). The corresponding rates of freedom from these cardiac events at 3 years were 93% ± 7%, 88% ± 8%, and 56% ± 17% ($P < .001$; Figure 2). The severity of preoperative PH was significantly associated with overall survival and the freedom from adverse cardiac events.

Acute Hemodynamic Changes After RMA

Preoperatively, all patients showed impairment of systemic and pulmonary hemodynamic conditions, the severity of which increased with the severity of the PH (Table 2, Figure 3). The systolic and mean PAP and PVR in the severe PH group were significantly greater than those in the other 2 groups.

From baseline to 1 month after surgery, the patients in the mild PH group showed good functional improvement, but those in the severe PH group showed less improvement. The patients in the moderate PH group showed improvement that was intermediate between that of the mild and severe PH groups. The LV volumes significantly decreased, and the LVEF improved in all 3 groups, although patients in the severe PH group showed less improvement in those parameters than in the other 2 groups. The LV systolic pressures did not change; however, the LV end-diastolic pressure decreased significantly in all groups.

In the patients with mild and moderate PH, the cardiac index increased significantly, the systemic vascular resistance decreased, and the systolic and mean PAP and PVR remained unchanged. In contrast, in the severe PH group, the cardiac index and systemic vascular resistance remained unchanged, and the systolic and mean PAP and the PVR significantly decreased. In the severe PH group, both systolic and mean PAP and PVR significantly decreased; Importantly, in patients with severe PH, the systolic PAP and PVR at 1 month after surgery were significantly greater than those in the mild PH group and remained substantially abnormal, suggesting the persistence of abnormal pulmonary hemodynamics.

Serial Echocardiographic Examinations

LV dimensions and function and left atrial dimensions. In the mild and moderate groups, the LVEDD, left ventricular end-systolic dimension, and left atrial dimension had decreased significantly and the LVEF had improved at 1 month after surgery (Table 3, Figure 4). Also, these improvements (reverse remodeling) persisted during the follow-up period. In the severe PH group, the LV dimensions remained unchanged at 1 month after surgery and were significantly greater than the other 2 groups. In addition, the severe PH group had lower LVEF values than the mild and moderate PH groups at all follow-up examinations.

Doppler-derived systolic PAP. In the mild and moderate PH groups, the mean systolic PAP at 1 month after surgery had remained or returned to a normal range in most patients. Those values had stabilized within the normal range during the follow-up period, along with an improvement in MR. In contrast, the mean systolic PAP in the severe PH group had decreased at 1 month after surgery; however, the values had never returned to a normal range, regardless of significant improvement in MR. The systolic PAP in the severe PH group had gradually worsened for a period of years, in contrast to the LV systolic function and LV dimension.

Mitral valve performance and measurements. Serial examinations showed significant improvements in MR in all groups, with optimal mitral valve geometry in terms of