62. Previtali M, Lanzarini L, Ferrario M, Tortorici M, Mussini A, Montemartini C. Dobutamine versus dipyridamole echocardiography in coronary artery disease. Circulation. 1991; 83(5 Suppl): III27–III31.
63. Sawada SG, Segar DS, Ryan T, et al. Echocardiographic detection of coronary artery disease during dobutamine infusion. Circulation 1991; 83(5): 1605–1614.
64. Epstein M, Gin K, Sterns L, Pollick C. Dobutamine stress echocardiography: initial experience of a Canadian centre. Can J Cardiol 1992; 8(3): 273–279.
65. Marcovitz PA, Armstrong WF. Accuracy of dobutamine stress echocardiography in detecting coronary artery disease. Am J Cardiol 1992; 69(16): 1269–1273.
66. Mazeika PK, Nadazdin A, Oakley CM. Dobutamine stress echocardiography for detection and assessment of coronary artery disease. J Am Coll Cardiol 1992; 19(6): 1203–1211.
67. McNeill AJ, Fioretti PM, el-Said SM, Salustri A, Forster T, Roelandt JR. Enhanced sensitivity for detection of coronary artery disease by addition of atropine to dobutamine stress echocardiography. Am J Cardiol 1992; 70(1): 41–46.
68. Salustri A, Fioretti PM, Pozzoli MM, McNeill AJ, Roelandt JR. Dobutamine stress echocardiography: its role in the diagnosis of coronary artery disease. Eur Heart J 1992; 13(1): 70–77.
69. Salustri A, Fioretti PM, McNeill AJ, Pozzoli MM, Roelandt JR. Pharmacological stress echocardiography in the diagnosis of coronary artery disease and myocardial ischaemia: a comparison between dobutamine and dipyridamole. Eur Heart J 1992; 13(10): 1356–1362.
70. Segar DS, Brown SE, Sawada SG, Ryan T, Feigenbaum H. Dobutamine stress echocardiography: correlation with coronary lesion severity as determined by quantitative angiography. J Am Coll Cardiol 1992; 19(6): 1197–1202.
71. Forster T, McNeill AJ, Salustri A, et al. Simultaneous dobutamine stress echocardiography and technetium-99m isonitrile single-photon emission computed tomography in patients with suspected coronary artery disease. J Am Coll Cardiol 1993; 21(7): 1591–1596.
72. Gunalp B, Dokumaci B, Uyan C, et al. Value of dobutamine technetium-99m-sestamibi SPECT and echocardiography in the detection of coronary artery disease compared with coronary angiography. J Nucl Med 1993; 34(6): 889–894.
73. Marwick T, D'Hondt AM, Baudhuin T, et al. Optimal use of dobutamine stress for the detection and evaluation of coronary artery disease: combination with echocardiography or scintigraphy, or both? J Am Coll Cardiol 1993; 22(1): 159–167.
74. Mazeika PK, Nadazdin A, Oakley CM. Prognostic value of dobutamine echocardiography in patients with high pretest likelihood of coronary artery disease. Am J Cardiol 1993; 71(1): 33–39.
75. Previtali M, Lanzarini L, Fetiveau R, et al. Comparison of dobutamine stress echocardiography, dipyridamole stress echocardiography and exercise stress testing for diagnosis of coronary artery disease. Am J Cardiol 1993; 72(12): 865–870.
76. Takeuchi M, Araki M, Nakashima Y, Kuroiwa A. Comparison of dobutamine stress echocardiography and stress thallium-201 single-photon emission computed tomography for detecting coronary artery disease. J Am Soc Echocardiogr 1993; 6(6): 593–602.
77. Warner MF, Pippin JJ, DiSciascio G, et al. Assessment of thallium scintigraphy and echocardiography during dobutamine infusion for the detection of coronary artery disease. Cathet Cardiovasc Diagn 1993; 29(2): 122–127.
78. Mairesse GH, Marwick TH, Vanoverschelde JL, et al. How accurate is dobutamine stress electrocardiography for detection of coronary artery disease? Comparison with two-dimensional echocardiography and technetium-99m methoxyl isobutyl isonitrile (mibi) perfusion scintigraphy. J Am Coll Cardiol 1994; 24(4): 920–927.
79. Ostojic M, Picano E, Beleslin B, et al. Dipyridamole-dobutamine echocardiography: a novel test for the detection of milder forms of coronary artery disease. J Am Coll Cardiol 1994; 23(5): 1115–1122.
80. Panza JA, Laurienzo JM, Curiel RV, Quyyumi AA, Cannon RO, 3rd. Transesophageal dobutamine stress echocardiography for evaluation of patients with coronary artery disease. J Am Coll Cardiol 1994; 24(5): 1260–1267.
81. Prince CR, Stoddard MF, Morris GT, et al. Dobutamine two-dimensional transesophageal echocardiographic stress testing for detection of coronary artery disease. Am Heart J 1994; 128(1): 36–41.
82. Sahin M, Karakelleoglu S, Alp N, Atesal S, Senocak H. Diagnostic value of dobutamine stress echocardiography in coronary artery disease. Thorac Cardiovasc Surg 1994; 42(5): 285–289.
83. Santiago P, Vacek JL, Rosamond TL. Dobutamine stress echocardiography: clinical utility and predictive value at various infusion rates. Am Heart J 1994; 128(4): 804–808.
84. Senior R, Sridhara BS, Anagnostou E, Handler C, Raftery EB, Lahiri A. Synergistic value of simultaneous stress dobutamine sestamibi single-photon-emission computerized tomography and echocardiography in the detection of coronary artery disease. Am Heart J 1994; 128(4): 713–718.
85. Castini D, Gentile F, Ornaghi M, et al. Dobutamine echocardiography: usefulness of digital image processing. Eur Heart J 1995; 16(10): 1420–1424.
86. Chan RK, Tonkin AM, Byrgiotis S, Calafiore P. Assessment of coronary artery disease by dobutamine stress echocardiography (DSE). Aust N Z J Med 1995; 25(6): 707–715.
87. Daoud EG, Pitt A, Armstrong WF. Electrocardiographic response during dobutamine stress echocardiography. Am Heart J 1995; 129(4): 672–677.
88. Frohwein S, Klein JL, Lane A, Taylor WR. Transesophageal dobutamine stress echocardiography in the evaluation of coronary artery disease. J Am Coll Cardiol 1995; 25(4): 823–829.
89. Geleijnse ML, Marwick TH, Boersma E, Deckers JW, Melin JA, Fioretti PM. Optimal pharmacological stress testing for the diagnosis of coronary artery disease: a

probabilistic approach. Eur Heart J 1995; 16(Suppl M): 3–10.
90. Latcham AP, Orsinelli DA, Pearson AC. Recognition of the segmental tendency of false-positive dobutamine stress echocardiograms and its effects on test sensitivity and specificity. Am Heart J 1995; 129(5): 1047–1050.
91. Mairesse GH, Marwick TH, Arnese M, et al. Improved identification of coronary artery disease in patients with left bundle branch block by use of dobutamine stress echocardiography and comparison with myocardial perfusion tomography. Am J Cardiol 1995; 76(5): 321–325.
92. Reis G, Marcovitz PA, Leichtman AB, et al. Usefulness of dobutamine stress echocardiography in detecting coronary artery disease in end-stage renal disease. Am J Cardiol 1995; 75(10): 707–710.
93. Sochowski RA, Yvorchuk KJ, Yang Y, Rattes MF, Chan KL. Dobutamine and dipyridamole stress echocardiography in patients with a low incidence of severe coronary artery disease. J Am Soc Echocardiogr 1995; 8(4): 482–487.
94. De Bello V, Bellina CR, Molea N, et al. Simultaneous dobutamine stress echocardiography and dobutamine scintigraphy (99mTc-MIBI-SPET) for assessment of coronary artery disease. Int J Card Imaging 1996; 12(3): 185–190.
95. Elhendy A, Geleijnse L, Salustri A, et al. T wave normalization during dobutamine stress testing in patients with non-Q wave myocardial infarction. A marker of myocardial ischaemia? Eur Heart J 1996; 17(4): 526–531.
96. Elhendy A, yan Domburg RT, Roelandt JR, et al. Accuracy of dobutamine stress echocardiography for the diagnosis of coronary artery stenosis in patients with myocardial infarction: the impact of extent and severity of left ventricular dysfunction. Heart 1996; 76(2): 123–128.
97. Hoffmann R, Lethen H, Marwick T, et al. Analysis of interinstitutional observer agreement in interpretation of dobutamine stress echocardiograms. J Am Coll Cardiol 1996; 27(2): 330–336.
98. Iwase M, Fukui M, Tamagaki H, et al. Advantages and disadvantages of dobutamine stress echocardiography compared with treadmill exercise electrocardiography in detecting ischemia. Jpn Circ J 1996; 60(12): 954–960.
99. Pingitore A, Picano E, Colosso MQ, et al. The atropine factor in pharmacologic stress echocardiography. Echo Persantine (EPIC) and Echo Dobutamine International Cooperative (EDIC) Study Groups. J Am Coll Cardiol 1996; 27(5): 1164–1170.
100. San Roman JA, Vilacosta I, Castillo JA, et al. Dipyridamole and dobutamine-atropine stress echocardiography in the diagnosis of coronary artery disease. Comparison with exercise stress test, analysis of agreement, and impact of antianginal treatment. Chest 1996; 110(5): 1248–1254.
101. Senior R, Basu S, Handler C, Raftery EB, Lahiri A. Diagnostic accuracy of dobutamine stress echocardiography for detection of coronary heart disease in hypertensive patients. Eur Heart J 1996; 17(2): 289–295.
102. Takeuchi M, Sonoda S, Miura Y, Kuroiwa A. Comparative diagnostic value of dobutamine stress echocardiography and stress thallium-201 single-photon-emission computed tomography for detecting coronary artery disease in women. Coron Artery Dis 1996; 7(11): 831–835.
103. Yeo TC, Ling LH, Ng WL, et al. Dobutamine stress echocardiography: angiographic correlates. Ann Acad Med Singapore 1996; 25(2): 196–199.
104. Hennessy TG, Codd MB, Kane G, McCarthy C, McCann HA, Sugrue DD. Dobutamine stress echocardiography in the detection of coronary artery disease: importance of the pretest likelihood of disease. Am Heart J 1997; 134(4): 685–692.
105. Hennessy TG, Codd MB, Kane G. Evaluation of patients with diabetes mellitus for coronary artery disease using dobutamine stress echocardiography. Coron Artery Dis 1997; 8(3–4): 171–174.
106. Ho YL, Wu CC, Huang PJ, et al. Dobutamine stress echocardiography compared with exercise thallium-201 single-photon emission computed tomography in detecting coronary artery disease-effect of exercise level on accuracy. Cardiology 1997; 88(4): 379–385.
107. Ho YL, Wu CC, Lin LC, et al. Assessment of the functional significance of coronary artery stenosis by dobutamine-atropine stress echocardiography. Cardiology 1997; 88(4): 386–392.
108. Oguzhan A, Kisacik HL, Ozdemir K, et al. Comparison of exercise stress testing with dobutamine stress echocardiography and exercise technetium-99m isonitrile single photon emission computerized tomography for diagnosis of coronary artery disease. Jpn Heart J 1997; 38(3): 333–344.
109. Vitarelli A, Luzzi MF, Penco M, Fedele F, Dagianti A. On-line quantitative assessment of left ventricular filling during dobutamine stress echocardiography: a useful addition to conventional wall motion scoring. Int J Cardiol 1997; 59(1): 57–69.
110. Elhendy A, van Domburg RT, Bax JJ, et al. Optimal criteria for the diagnosis of coronary artery disease by dobutamine stress echocardiography. Am J Cardiol 1998; 82(11): 1339–1344.
111. Elhendy A, Geleijnse ML, van Domburg RT, et al. Comparison of dobutamine stress echocardiography and technetium-99m sestamibi single-photon emission tomography for the diagnosis of coronary artery disease in hypertensive patients with and without left ventricular hypertrophy. Eur J Nucl Med 1998; 25(1): 69–78.
112. Elhendy A, van Domburg RT, Poldermans D, et al. Safety and feasibility of dobutamine-atropine stress echocardiography for the diagnosis of coronary artery disease in diabetic patients unable to perform an exercise stress test. Diabetes Care 1998; 21(11): 1797–1802.
113. Elhendy A, van Domburg RT, Bax JJ, et al. Noninvasive diagnosis of coronary artery stenosis in women with limited exercise capacity: comparison of dobutamine stress echocardiography and 99mTc sestamibi single-photon emission CT. Chest 1998; 114(4): 1097–1104.
114. Hennessy TG, Siobhan Hennessy M, Codd MB, et al. Detection of coronary artery disease using dobutamine stress echocardiography in patients with an abnormal resting electrocardiograph. Int J Cardiol 1998; 64(3): 293–298.

115. Ho YL, Wu CC, Huang PJ, et al. Assessment of coronary artery disease in women by dobutamine stress echocardiography: comparison with stress thallium-201 single-photon emission computed tomography and exercise electrocardiography. Am Heart J 1998; 135(4): 655–662.
116. Ho YL, Wu CC, Lin LC, et al. Assessment of the coronary artery disease and systolic dysfunction in hypertensive patients with the dobutamine-atropine stress echocardiography: effect of the left ventricular hypertrophy. Cardiology 1998; 89(1): 52–58.
117. Khattar RS, Senior R, Lahiri A. Assessment of myocardial perfusion and contractile function by inotropic stress Tc-99m sestamibi SPECT imaging and echocardiography for optimal detection of multivessel coronary artery disease. Heart 1998; 79(3): 274–280.
118. San Roman JA, Vilacosta I, Castillo JA, et al. Selection of the optimal stress test for the diagnosis of coronary artery disease. Heart 1998; 80(4): 370–376.
119. Santoro GM, Sciagra R, Buonamici P, et al. Head-to-head comparison of exercise stress testing, pharmacologic stress echocardiography, and perfusion tomography as first-line examination for chest pain in patients without history of coronary artery disease. J Nucl Cardiol 1998; 5(1): 19–27.
120. Shaheen J, Luria D, Klutstein MW, Rosenmann D, Tzivoni D. Diagnostic value of 12-lead electrocardiogram during dobutamine echocardiographic studies. Am Heart J 1998; 136(6): 1061–1064.
121. Elhendy A, Bax JJ, van Domburg RT, et al. Dobutamine stress thallium-201 single-photon emission tomography versus echocardiography for evaluation of the extent and location of coronary artery disease late after myocardial infarction. Eur J Nucl Med 1999; 26(5): 467–473.
122. Fragasso G, Lu C, Dabrowski P, Pagnotta P, Sheiban I, Chierchia SL. Comparison of stress/rest myocardial perfusion tomography, dipyridamole and dobutamine stress echocardiography for the detection of coronary disease in hypertensive patients with chest pain and positive exercise test. J Am Coll Cardiol 1999; 34(2): 441–447.
123. Herzog CA, Marwick TH, Pheley AM, White CW, Rao VK, Dick CD. Dobutamine stress echocardiography for the detection of significant coronary artery disease in renal transplant candidates. Am J Kidney Dis 1999; 33(6): 1080–1090.
124. Hoffmann R, Lethen H, Kuhl H, Lepper W, Hanrath P. Extent and severity of test positivity during dobutamine stress echocardiography. Influence on the predictive value for coronary artery disease. Eur Heart J 1999; 20(20): 1485–1492.
125. Nagel E, Lehmkuhl HB, Bocksch W, et al. Noninvasive diagnosis of ischemia-induced wall motion abnormalities with the use of high-dose dobutamine stress MRI: comparison with dobutamine stress echocardiography. Circulation 1999; 99(6): 763–770.
126. Ozdemir K, Kisacik HL, Oguzhan A, et al. Comparison of exercise stress testing with dobutamine stress echocardiography and radionuclide ventriculography for diagnosis of coronary artery disease. Jpn Heart J 1999; 40(6): 715–727.
127. Therre T, Ribal JP, Motreff P, et al. Assessment of cardiac risk before aortic reconstruction: noninvasive work-up using clinical examination, exercise testing, and dobutamine stress echocardiography versus routine coronary arteriography. Ann Vasc Surg 1999; 13(5): 501–508.
128. Ariff B, Thom SA, Foale RA, Mayet J. Stress echocardiography for the diagnosis of ischaemia in hypertensives. J Hum Hypertens 2000; 14(6): 399–401.
129. Joseph T, Vieillard-Baron A, Chikli F, et al. Left ventricular volume analysis for the detection of coronary artery disease during dobutamine stress echocardiography in patients undergoing vascular surgery. Eur J Echocardiogr 2000; 1(4): 263–270.
130. Smart SC, Knickelbine T, Malik F, Sagar KB. Dobutamine-atropine stress echocardiography for the detection of coronary artery disease in patients with left ventricular hypertrophy. Importance of chamber size and systolic wall stress. Circulation 2000; 101(3): 258–263.
131. Ahmad M, Xie T, McCulloch M, Abreo G, Runge M. Real-time three-dimensional dobutamine stress echocardiography in assessment stress echocardiography in assessment of ischemia: comparison with two-dimensional dobutamine stress echocardiography. J Am Coll Cardiol 2001; 37(5): 1303–1309.
132. Dolan MS, Riad K, El-Shafei A, et al. Effect of intravenous contrast for left ventricular opacification and border definition on sensitivity and specificity of dobutamine stress echocardiography compared with coronary angiography in technically difficult patients. Am Heart J 2001; 142(5): 908–915.
133. Peteiro J, Monserrat L, Fabregas R, Manuel Vazquez J, Calvino R, Castro-Beiras A. Comparison of two-dimensional echocardiography and pulsed Doppler tissue imaging during dobutamine-atropine stress testing to detect coronary artery disease. Echocardiography 2001; 18(4): 275–284.
134. Picano E, Distante A, Masini M, Morales MA, Lattanzi F, L'Abbate A. Dipyridamole-echocardiography test in effort angina pectoris. Am J Cardiol 1985; 56(7): 452–456.
135. Picano E, Lattanzi F, Masini M, Distante A, L'Abbate A. High dose dipyridamole echocardiography test in effort angina pectoris. J Am Coll Cardiol 1986; 8(4): 848–854.
136. Picano E, Masini M, Lattanzi F, Distante A, L'Abbate A. Role of dipyridamole-echocardiography test in electrocardiographically silent effort myocardial ischemia. Am J Cardiol 1986; 58(3): 235–237.
137. Picano E, Morales MA, Distante A, et al. Dipyridamole-echocardiography test in angina at rest: noninvasive assessment of coronary stenosis underlying spasm. Am Heart J 1986; 111(4): 688–691.
138. Picano E, Lattanzi F, Masini M, Distante A, L'Abbate A. Comparison of the high-dose dipyridamole-echocardiography test and exercise two-dimensional echocardiography for diagnosis of coronary artery disease. Am J Cardiol 1987; 59(6): 539–542.
139. Masini M, Picano E, Lattanzi F, Distante A, L'Abbate A. High dose dipyridamole-echocardiography test in women:

correlation with exercise-electrocardiography test and coronary arteriography. J Am Coll Cardiol 1988; 12(3): 682–685.
140. Picano E, Masini M, Lattanzi F, et al. Short term reproducibility of exercise testing in patients with ST segment elevation and different responses to the dipyridamole test. Br Heart J 1988; 60(4): 281–286.
141. Picano E, Lattanzi F, Masini M, Distante A, L'Abbate A. Usefulness of the dipyridamole-exercise echocardiography test for diagnosis of coronary artery disease. Am J Cardiol 1988; 62(1): 67–70.
142. Picano E, Severi S, Michelassi C, et al. Prognostic importance of dipyridamole-echocardiography test in coronary artery disease. Circulation 1989; 80(3): 450–457.
143. Ferrara N, Leosco D, Abete P, et al. Dipyridamole echocardiography as a useful and safe test in the assessment of coronary artery disease in the elderly. J Am Geriatr Soc 1991; 39(10): 993–999.
144. Mandysova E, Niederle P, Malkova A, et al. Usefulness of dipyridamole-handgrip echocardiography test for detecting coronary artery disease. Am J Cardiol 1991; 67(9): 883–885.
145. Perin EC, Moore W, Blume M, Hernandez G, Dhekne R, DeCastro CM. Comparison of dipyridamole-echocardiography with dipyridamole-thallium scintigraphy for the diagnosis of myocardial ischemia. Clin Nucl Med 1991; 16(6): 417–420.
146. Picano E, Lattanzi F, Orlandini A, Marini C, L'Abbate A. Stress echocardiography and the human factor: the importance of being expert. J Am Coll Cardiol 1991; 17(3): 666–669.
147. Agati L, Renzi M, Sciomer S, et al. Transesophageal dipyridamole echocardiography for diagnosis of coronary artery disease. J Am Coll Cardiol 1992; 19(4): 765–770.
148. Lattanzi F, Picano E, Frugoli A, et al. Oral vs intravenous dipyridamole echocardiography for detecting coronary artery disease. Chest 1992; 102(4): 1189–1192.
149. Mazeika P, Nihoyannopoulos P, Joshi J, Oakley CM. Uses and limitations of high dose dipyridamole stress echocardiography for evaluation of coronary artery disease. Br Heart J 1992; 67(2): 144–149.
150. Bjoernstad K, Aakhus S, Lundbom J, et al. Digital dipyridamole stress echocardiography in silent ischemia after coronary artery bypass grafting and/or after healing of acute myocardial infarction. Am J Cardiol 1993; 72(9): 640–646.
151. Picano E, Pingitore A, Conti U, et al. Enhanced sensitivity for detection of coronary artery disease by addition of atropine to dipyridamole echocardiography. Eur Heart J 1993; 14(9): 1216–1222.
152. Severi S, Picano E, Michelassi C, et al. Diagnostic and prognostic value of dipyridamole echocardiography in patients with suspected coronary artery disease. Comparison with exercise electrocardiography. Circulation 1994; 89(3): 1160–1173.
153. Bjornstad K, Aakhus S, Hatle L. Digital high frame rate stress echocardiography for detection of coronary artery stenosis by high dose dipyridamole stress testing. Int J Card Imaging 1995; 11(3): 163–170.
154. Lanzarini L, Fetiveau R, Poli A, Diotallevi P, Barberis P, Previtali M. Results of dipyridamole plus atropine echo stress test for the diagnosis of coronary artery disease. Int J Card Imaging 1995; 11(4): 233–240.
155. Wagdi P, Kaufmann U, Fluri M, Meier B. High dose dipyridamole as a pharmacological stress test during cardiac catheterisation in patients with coronary artery disease. Heart 1996; 75(3): 247–251.
156. Bjornstad K, Aakhus S, Torp HG. How does computer-assisted digital wall motion analysis influence observer agreement and diagnostic accuracy during stress echocardiography? Int J Card Imaging 1997; 13(2): 105–114.
157. Gaddi O, Tortorella G, Picano E, et al. Diagnostic and prognostic value of vasodilator stress echocardiography in asymptomatic Type 2 diabetic patients with positive exercise thallium scintigraphy: a pilot study. Diabet Med 1999; 16(9): 762–766.
158. Parodi G, Picano E, Marcassa C, et al. High dose dipyridamole myocardial imaging: simultaneous sestamibi scintigraphy and two-dimensional echocardiography in the detection and evaluation of coronary artery disease. Italian Group of Nuclear Cardiology. Coron Artery Dis 1999; 10(3): 177–184.
159. Astarita C, Palinkas A, Nicolai E, Maresca FS, Varga A, Picano E. Dipyridamole-atropine stress echocardiography versus exercise SPECT scintigraphy for detection of coronary artery disease in hypertensives with positive exercise test. J Hypertens 2001; 19(3): 495–502.
160. Zabalgoitia M, Gandhi DK, Abi-Mansour P, Yarnold PR, Moushmoush B, Rosenblum J. Transesophageal stress echocardiography: detection of coronary artery disease in patients with normal resting left ventricular contractility. Am Heart J 1991; 122(5): 1456–1463.
161. de Cock CC, Kamp O, Meijer A. Identifying patients for rate responsive atrial pacing: a new method for patient selection and pacemaker programming. Pacing Clin Electrophysiol 1992; 15(11 Pt 2): 1792–1797.
162. Kamp O, De Cock CC, Kupper AJ, Roos JP, Visser CA. Simultaneous transesophageal two-dimensional echocardiography and atrial pacing for detecting coronary artery disease. Am J Cardiol 1992; 69(17): 1412–1416.
163. Norris LP, Stewart RE, Jain A, Hibner CS, Chaudhuri TK, Zabalgoitia M. Biplane transesophageal pacing echocardiography compared with dipyridamole thallium-201 single-photon emission computed tomography in detecting coronary artery disease. Am Heart J 1993; 126(3 Pt 1): 676–685.
164. Hoffmann R, Kleinhans E, Lambertz H, et al. Transoesophageal pacing echocardiography for detection of restenosis after percutaneous transluminal coronary angioplasty. Eur Heart J 1994; 15(6): 823–831.
165. Michael TA, Rao G, Balasingam S. Accuracy and usefulness of atrial pacing in conjunction with transesophageal echocardiography in the detection of cardiac ischemia (a comparative study with scintigraphic tomography and coronary arteriography). Am J Cardiol 1995; 75(8): 563–567.
166. Iliceto S, Sorino M, D'Ambrosio G, et al. Detection of coronary artery disease by two-dimensional echocardiography and transesophageal atrial pacing. J Am Coll Cardiol 1985; 5(5): 1188–1197.

167. Matthews RV, Haskell RJ, Ginzton LE, Laks MM. Usefulness of esophageal pill electrode atrial pacing with quantitative two-dimensional echocardiography for diagnosing coronary artery disease. Am J Cardiol 1989; 64(12): 730–735.
168. Laucevicius A, Berukstis E, Kosinskas E, Jablonskiene D, Ivaskeviciene L. Quantitative assessment of transient left ventricular asynergy in coronary artery disease. Clin Cardiol 1991; 14(2): 105–110.
169. Michael TA, Antonescu A, Bhambi B, Balasingam S. Accuracy and usefulness of atrial pacing in conjunction with transthoracic echocardiography in the detection of cardiac ischemia. Am J Cardiol 1996; 77(2): 187–190.
170. Atar S, Nagai T, Cercek B, Naqvi TZ, Luo H, Siegel RJ. Pacing stress echocardiography: an alternative to pharmacologic stress testing. J Am Coll Cardiol 2000; 36(6): 1935–1941.
171. O'Keefe J, Jr, Barnhart C, Bateman T. Comparison of stress echocardiography and stress myocardial perfusion scintigraphy for diagnosing coronary artery disease and assessing its severity. Am J Cardiol 1995; 75(11): 25D–34D.
172. Fleischmann KE, Hunink MG, Kuntz KM, Douglas PS. Exercise echocardiography or exercise SPECT imaging? A meta-analysis of diagnostic test performance. JAMA 1998; 280(10): 913–920.
173. Kwok Y, Kim C, Grady D, Segal M, Redberg R. Meta-analysis of exercise testing to detect coronary artery disease in women. Am J Cardiol 1999; 83(5): 660–666.
174. Picano E, Bedetti G, Varga A, Cseh E. The comparable diagnostic accuracies of dobutamine-stress and dipyridamole-stress echocardiographies: a meta-analysis. Coron Artery Dis 2000; 11(2): 151–159.
175. Kim C, Kwok YS, Heagerty P, Redberg R. Pharmacologic stress testing for coronary disease diagnosis: a meta-analysis. Am Heart J 2001; 142(6): 934–944.
176. Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. Ann Intern Med 1994; 120(8): 667–676.
177. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. N Engl J Med 1978; 299(17): 926–930.
178. Rozanski A, Diamond GA, Berman D, Forrester JS, Morris D, Swan HJ. The declining specificity of exercise radionuclide ventriculography. N Engl J Med 1983; 309(9): 518–522.

Address for correspondence: Yoshinori Noguchi, Department of General Internal Medicine, Fujita Health University School of Medicine, 1-98 Dengakugakubo, Kutsukake-cho, Toyoake-city, Aichi prefecture 470-1192, Japan
Tel.: +81-0562-93-2000; Fax: +81-0562-93-3797
E-mail: yoshi-noguchi@umin.ac.jp

Ventilation-Perfusion Scanning and Helical CT in Suspected Pulmonary Embolism: Meta-Analysis of Diagnostic Performance[1]

Yasuaki Hayashino, MD, MPH
Masashi Goto, MD, MPH
Yoshinori Noguchi, MD, MPH
Tsuguya Fukui, MD, MPH

Published online
10.1148/radiol.2343031009
Radiology 2005; 234:740–748

Abbreviations:
CI = confidence interval
MAA = macroaggregated albumin
PE = pulmonary embolism
PIOPED = Prospective Investigation of Pulmonary Embolism Diagnosis
PYP = pyrophosphate
ROC = receiving operator characteristic
V-P = ventilation-perfusion

[1] From the Department of General Medicine and Clinical Epidemiology, Kyoto University Graduate School of Medicine, Kyoto, Japan. Received June 26, 2003; revision requested September 9; final revision received May 21, 2004; accepted June 23. **Address correspondence to** T.F. Department of Medicine, St Luke's International Hospital, 9–1 Akashi-cho, Chuo-ku, Tokyo, 104-8560, Japan. (e-mail: *fkts@luke.or.jp*).

Authors stated no financial relationship to disclose.

Author contributions:
Guarantor of integrity of entire study, Y.H.; study concepts and design, all authors; literature research, Y.H., M.G.; data acquisition, Y.H., G.M.; data analysis/interpretation, all authors; statistical analysis, Y.H., Y.N.; manuscript preparation and definition of intellectual content, all authors; manuscript editing and revision/review, Y.H., Y.N., T.F.; manuscript final version approval, all authors

PURPOSE: To perform meta-analysis of literature about the role of helical computed tomography (CT) and ventilation-perfusion (V-P) scanning in detection of acute pulmonary embolism (PE) by using summary receiver operating characteristic (ROC) curve analysis.

MATERIALS AND METHODS: V-P scanning articles published from January 1985 to March 2003 and helical CT articles published from January 1990 to March 2003 in MEDLINE and EMBASE databases were included if *(a)* tests were performed for evaluation of acute PE, *(b)* conventional angiography was the reference standard, and *(c)* absolute numbers of true-positive, false-negative, true-negative, and false-positive results were available. Sensitivity analysis was conducted by excluding articles published before 1995.

RESULTS: A total of 12 articles discussing helical CT and/or V-P scanning were included. With a random-effects model, pooled sensitivity for helical CT was 86.0% (95% confidence interval [CI]: 80.2%, 92.1%), and specificity was 93.7% (95% CI: 91.1%, 96.3%). V-P scanning yielded low sensitivity of 39.0% (95% CI: 37.3%, 40.8%) but high specificity of 97.1% (95% CI: 96.0%, 98.3%) with high probability threshold. V-P scanning yielded high sensitivity of 98.3% (95% CI: 97.2%, 99.5%) and low specificity of 4.8% (95% CI: 4.7%, 4.9%) with normal threshold. Regression coefficients for helical CT angiography were 0.588 (95% CI: −1.55, 2.74) and 4.14 (95% CI: −0.002, 8.28) versus V-P scanning with high and normal thresholds, respectively. Regression coefficients for helical CT angiography were 0.588 (95% CI: −1.55, 2.74) and 4.14 (95% CI: −0.002, 8.28) versus V-P scanning with high and normal thresholds, respectively.

CONCLUSION: Helical CT has greater discriminatory power than V-P scanning with normal and/or near-normal threshold to exclude PE, while helical CT and V-P scanning with high probability threshold had similar discriminatory power in the diagnosis of PE.

The diagnosis of pulmonary embolism (PE) remains difficult because the symptoms are not specific and because all available tests have substantial limitations in clinical settings (1,2). Selective pulmonary angiography remains the reference standard; however, this procedure is invasive and causes morbidity and mortality rates of 4% and 0.2%, respectively (3,4). Moreover, it is costly and time-consuming. Several minimally invasive modalities have been used to facilitate detection of PE. Because of its general availability, ventilation-perfusion (V-P) scanning is one of the most frequently used procedures in the confirmation or exclusion of PE. The Prospective Investigation of Pulmonary Embolism Diagnosis (PIOPED) study (5) assessed the usefulness of V-P lung scanning in patients with acute PE. Despite accuracy rates as high as 96% in some instances, V-P scanning alone is inadequate

to confirm or exclude diagnosis of PE because up to 75% of patients are in the inconclusive category; therefore, other modalities, such as serial compression ultrasonography and D-dimer testing, are useful adjuncts (6,7).

Contrast agent–enhanced helical computed tomography (CT) of the pulmonary arteries has been proposed, and data are accumulating (8–17). The choice between V-P scanning and helical CT should be determined by using available data to compare diagnostic accuracy. Interpretation of the data as a whole is difficult because of the wide variation in the background of the patients. In prior reports that compare the test performance of helical CT and V-P scanning (18,19), the difference in test performance has not been systematically compared with a statistical method frequently used in meta-analysis of diagnostic tests.

Summary receiver operating characteristic (ROC) analysis is a method that enables quantitative combination of the multiple studies with heterogeneous results; each point on the summary ROC curve represents a combination of sensitivity and specificity that could result from each study (20–22). Heterogeneity among different studies is caused either by differences between the way clinicians define a test as positive for PE or by wide variation in terms of the patients' background; summary ROC curve analysis could resolve those problems by means of stringent inclusion criteria and meta-regression analysis. It is suitable to compare the performance of different diagnostic tests, and reports involving the use of this analysis have been accumulating (23,24). Thus, the purpose of our study was to perform meta-analysis of the helical CT and V-P scanning literature by using the methodologic tool of summary ROC analysis.

MATERIALS AND METHODS

Identification of Trials

A computerized search was performed to identify relevant English-language articles published in MEDLINE. To search for articles about helical CT, we used "computed tomography" in the text, "image interpretation, computer-assisted" in medical subject headings, and "pulmonary embolism" in medical subject headings or text. Similarly, we searched for articles about V-P scanning by using "radionuclide imaging" in medical subject headings and "perfusion scan" or "ventilation-perfusion lung scan" in the text. The search for helical CT articles was limited to articles published between January 1, 1990, and May 30, 2003, since earlier CT equipment was substantially different. We included articles on V-P scanning published from January 1, 1985, to May 30, 2003, since this technique was used in the diagnosis of PE beginning in the mid-1980s. We also scanned references in retrieved articles and contacted individuals who are knowledgeable in radiology to see if they knew of any other relevant reports. A database search with a similar strategy was conducted in the EMBASE database for articles published between 1985 and 2003. We did not search for unpublished reports.

Inclusion Criteria

We included a study if *(a)* helical CT or V-P scanning was used as a diagnostic tool for acute PE; *(b)* absolute numbers of true-positive, false-positive, true-negative, and false-negative cases or their equivalent were given; *(c)* pulmonary angiography was used as the reference standard for diagnosis of PE; and *(d)* the time interval between the findings obtained from the test and reference standard was 48 hours or less, taking into account the fact that PE might disappear during the interval of two tests. We determined this time interval by reviewing literature of diagnostic tests in patients with PE (25,26) and discussing this issue with radiologists.

A study was excluded if *(a)* pulmonary angiography, in combination with any other modality, served as the reference standard; *(b)* helical CT was not performed for acute PE (eg, chronic PE or septic embolism); *(c)* noncomparable CT methods (eg, electron-beam CT) were used; *(d)* helical CT was performed after anticoagulant therapy or surgery for PE; or *(e)* the published information was incomplete. We tried to see if different studies from the same institution used the same patients because one author published several reports.

Data Collection

Two investigators (Y.H., M.G.) independently abstracted the data from all articles included in our analysis. The information abstracted included descriptive data (eg, authors, title, journal citation, and year of publication), study group characteristics (eg, sample size, mean age, proportion of women, and prevalence of PE), study design characteristics (involving criteria used to define a positive result and protocol information), extent of blinding between readers, information about the extent of the disease, and any evidence of verification bias and test interpretation bias.

For each study, the results were classified as true-positive, false-positive, true-negative, and false-negative. For V-P scanning, PIOPED criteria are generally used according to the probability of PE (eg, high, intermediate, low, or near normal or normal) (5). We specify three criteria for calculating sensitivity and specificity (Table 1). High-probability V-P scanning findings were positive, and others (eg, intermediate probability, low probability, or near normal and/or normal) were negative (threshold 1). High- and intermediate-probability V-P scanning findings were positive, and low-probability and normal and/or near-normal V-P scanning findings were negative (threshold 2). Normal and/or near-normal V-P scanning findings were negative, and others (eg, high probability, intermediate probability, and low probability) were positive (threshold 3). In one study, researchers developed and used original criteria that consisted of five categories; these five categories were reduced to four to match PIOPED criteria (27). As for helical CT, the presence or absence of PE, defined as an intraluminal filling defect or complete nonfilling of a pulmonary artery, was used as a criterion for determining positive or negative status.

From the articles in which investigators tabulated the results for different observers, we extracted data for the first observer, unless one observer was emphasized in the literature. When authors emphasized one result (eg, level of experience) in different observers, we extracted the most emphasized results. Any inconsistencies or controversies encountered in abstracted data were resolved with discussion and consensus.

Analysis and Statistics

The overall suitability of the pooled and summary ROC curve analysis was evaluated by using the Spearman correlation coefficient (22). We then checked heterogeneity separately for sensitivity and specificity. Since sensitivities for helical CT and specificities for V-P scanning threshold 3 were not homogeneous ($P = .006$ and $P < .001$, respectively), pooled sensitivity and specificity estimates were calculated by using a random-effects model that weighted each report according to its sample size (28).

TABLE 1
Comparison between Thresholds and Probabilities according to PIOPED Criteria

PIOPED Criteria and Probability	Threshold 1	Threshold 2	Threshold 3
High	Positive	Positive	Positive
Intermediate	Negative	Positive	Positive
Low	Negative	Negative	Positive
Near normal and/or normal	Negative	Negative	Negative

To estimate the summary ROC curve for helical CT and V-P scanning, we used a previously described method of variance-weighted least squares regression (20,22,23,28). On the basis of the 2 × 2 table constructed from each report, we made a logit transformation of the true-positive (eg, sensitivity) and false-positive (eg, 1 minus specificity) rates. Differences in the logit transformations (eg, measure of the observed discriminatory power of helical CT and V-P scanning) were then regressed on the sums of the logit transformations (eg, measure of the positivity threshold used to determine positive helical CT and V-P scanning results). Summary ROC curves for helical CT and V-P scanning were constructed with back transformation of the fitted line from the regression model. We weighted each study in the regression model by its variance with the following equation: [1/(true-positive + 0.5)] + [1/(false-positive + 0.5)] + [1/(false-negative + 0.5)] + [1/(true-negative + 0.5)] (20). We restricted the final summary ROC curves to the range of observed true-positive and false-positive rates.

Adjustment for clinical variables was accomplished by including them in the regression model. Inclusion of a dummy variable in the regression analysis for the type of diagnostic examination performed (eg, 1 for helical CT and 0 for V-P scanning) allows comparison of tests. The regression coefficient of this dummy variable is a measure of the difference in discriminatory power between the examinations. A positive regression coefficient implies increased discriminatory power for helical CT compared with V-P scanning, and a negative regression coefficient implies reduced discriminatory power. To avoid undefined values for diagnostic odds ratio, positivity criteria, and the variance that arises from zeroes of the true-positive, false-negative, true-negative, or false-positive values, 0.5 was added to that value (20).

We assessed the effect of publication year, mean age (55 years or younger vs older than 55 years), prevalence of PE, duration of tests (<24 hours vs <48 hours), study design (prospective vs retrospective), presence of interpretation bias, and presence of verification bias (eg, presence of verification bias vs no available information) in a combined model of helical CT and V-P scanning. We could not consider the effect of the extent of the disease in the model used to compare helical CT and V-P scanning, since information about the extent of the disease was not available in the literature that mainly dealt with V-P scanning (5,27,29).

We dichotomized some variables (eg, age, percentage of women, and duration between tests) at median. Because of the availability of data (eg, data on collimation were only available for helical CT) or missing data, the following variables were analyzed separately in each model: percentage of women included in the study (eg, ≤25% vs >25%), collimation (eg, 3 mm or thinner vs thicker than 3 mm), size of PE (eg, segmental vs subsegmental) for helical CT model, and type of radionuclide (eg, technetium 99m [^{99m}Tc] diethylenetriaminepentaacetic acid [DTPA] vs other types of radionuclides) used for V-P scanning. In the combined model, univariate analysis was performed to enable the effect of each clinical covariate to be assessed.

We added the factors that had a *P* value of less than .20 at univariate analysis into a multivariate regression model and used backward elimination to remove variables with a *P* value of more than .05. For the main aim of this study, a dummy variable for the type of diagnostic test (helical CT = 1) was always kept in this process. In separate and combined models, V-P scanning data were treated separately in different circumstances (eg, helical CT vs V-P scanning threshold 1, helical CT vs V-P scanning threshold 2, and helical CT vs V-P scanning threshold 3). Finally, we reanalyzed the final model with random-effects regression analysis (Technical bulletin no. 42; Stata Statistical Software, College Station, Tex), which took inter- and intrastudy variability into account.

After sensitivities and specificities were pooled, we assessed the posttest probability of PE on the basis of different pretest probabilities (low = .03, moderate = .27, high = .78). The arbitrary pretest probabilities of .03, .27, and .78 were based on the report by Wells et al (1), in which pretest probability was determined by using clinical signs and symptoms. First, pretest odds were converted into posttest odds by multiplying the pretest odds by the likelihood ratio. Likelihood ratio is defined as the probability of the test result in people with the disease divided by the probability of the test result in people without the disease. Posttest odds were converted back to posttest probabilities (30).

Since the helical CT method used in the detection of PE has undergone rapid changes in the past decade, the test performance might have changed from the early 1990s to 2003. Thus, we performed a sensitivity analysis by excluding helical CT articles that were published before 1995 and compared the results with those in base-case analysis. All analyses were performed by using commercially available software (Intercooled Stata 7.0; StataCorp, College Station, Tex).

RESULTS

Summary of the Literature Review and Data Extraction

Our initial data search yielded a total of 1385 titles of studies that used helical CT or V-P scanning. We excluded 1306 articles by reviewing titles and abstracts and selected 79 possible articles for our analysis; the reasons for exclusion are summarized in Figure 1. Of these 79 articles, we identified 12 that met all the inclusion criteria (5,8–11,13–17,27,29). Two of these articles also reported test performance of both helical CT and V-P scanning (10,14); thus, we had nine articles for helical CT and five articles for V-P scanning. Two studies were reported by the same lead author (Remy-Jardin) (8,11), but these studies were judged not to be overlapped because of the different study periods and different amount of contrast agent used. Overall, there were discrepancies between the two authors in 19 (13.6%) of 140 extracted items, ranging from 0% to 35.7% depending on the sort of items extracted. All discrepancies were resolved by consensus.

Helical CT and V-P Scanning

Helical CT was used to aid diagnosis of PE for the first time in 1992 (8). Of 12 articles we identified, nine collectively reported findings in 520 subjects (range,

10–151 subjects per study) who underwent helical CT (8–11,13–17), while five reported findings in 1269 subjects (range, 20–731 subjects) who underwent V-P scanning (5,10,14,27,29).

Variations in study protocols included thickness of the scanning section (range, 2.5–5.0 mm) and the amount of contrast agent (range, 70–150 mL). As for detector system, single-detector row helical CT was used in eight studies, and dual-detector row helical CT was used in one study (17). The reported sensitivity of helical CT ranged from 53% (13) to 100% (8,9), and specificity ranged from 75% (17) to 100% (9,11,14).

Ventilation studies used xenon 133 gas (^{133}Xe), ^{99m}Tc-pyrophosphate (PYP), or ^{99m}Tc-DTPA as the nuclear isotope, and perfusion studies used ^{99m}Tc-macroaggregated albumin (MAA) as the nuclear isotope. PIOPED criteria were used in four studies, and original criteria were formulated in one study (27). The reported specificity of V-P scanning with threshold 1 ranged from 96.0% (5) to 100% (10,14). With threshold 2, sensitivity ranged from 54.5% (10) to 100% (14), and with threshold 3, sensitivity ranged from 98% (5) to 100% (10,14,27,29). Details of the articles included are summarized in Table 2.

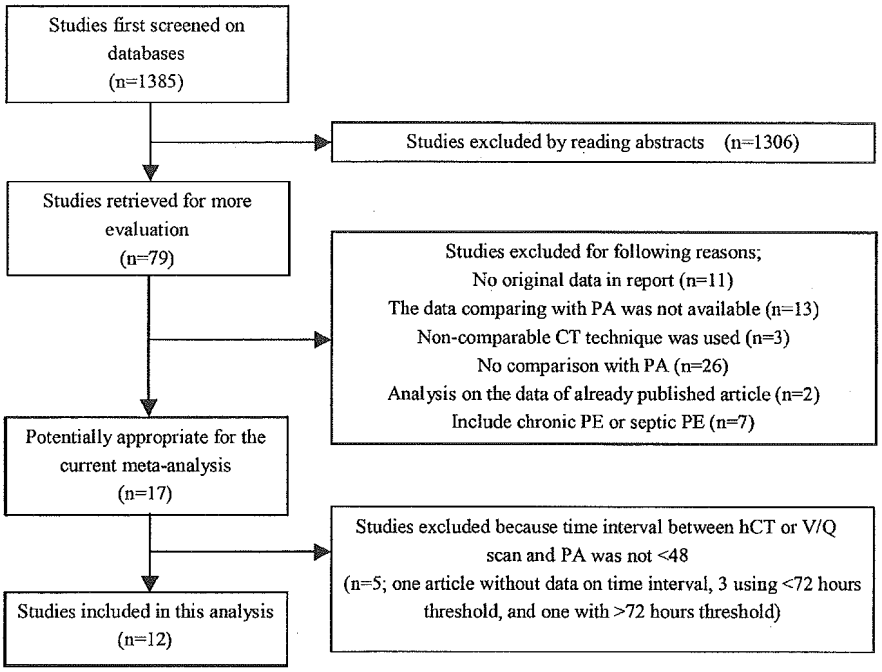

Figure 1. Flow chart shows results of literature search. *hCT* = helical CT, *PA* = pulmonary angiography.

Weighted Pooled Results

Weighted pooled data are presented in Table 3. Compared with helical CT, V-P scanning yielded a sensitivity of 39.0% (95% CI: 37.3%, 40.8%) with threshold 1, while it yielded sensitivities of 86.0% (95% CI: 83.3%, 88.8%) and 98.4% (95% CI: 97.2, 99.5) with thresholds 2 and 3, respectively. V-P scanning yielded a specificity of 97.1% (95% CI: 96.0%, 98.3%) with threshold 1 and a sensitivity of 39.0% (95% CI: 37.3%, 40.8%) and 4.8% (95% CI: 4.7%, 4.9%) with thresholds 2 and 3, respectively.

Summary ROC Analysis

No significant predictors were found in the separate univariate analysis for helical CT (eg, collimation or size of PE) or V-P scanning (eg, type of radionuclide used for V-P scanning). Univariate analysis for the comparison of two tests revealed that the following variables had a *P* value of less than .20: age in the model (P = .032), including V-P scanning threshold 1; duration between two tests (P = .027) and presence of verification bias (P = .067) in the model, including V-P scanning threshold 2; and duration between two tests (P = .111) and presence of verification bias (P = .146) in the model, including V-P scanning threshold 3. Thus, we included these variables in each multivariate model; however, as a result of backward elimination, no significant predictors were kept in the final model. In a final model used to compare helical CT and V-P scanning threshold 2, helical CT displayed superior discriminatory power, with a β coefficient of 3.73 (95% CI: 2.56, 4.9). In a model used to compare performance of helical CT with that of V-P scanning threshold 3, the estimate was large enough to suggest that helical CT was superior to V-P scanning, although this difference was not statistically significant (β coefficient, 4.14; P = .05). In the comparison of helical CT and V-P scanning threshold 1, no significant difference was observed in terms of discriminatory power (β coefficient, 0.588; P = .457). These results are shown in Table 4 and Figure 2.

Posttest Probability of PE

Posttest probabilities of PE according to positive or negative results of each test are summarized in Table 5. When pretest probabilities were moderate (.27) or high (.78), posttest probabilities of positive results of V-P scanning with threshold 1 and helical CT were .841 and .98, respectively. When pretest probability was low (.03), posttest probability of negative helical CT was .005, and that of negative V-P scanning with threshold 3 was .01. When pretest probabilities were moderate or high, posttest probabilities of negative helical CT and V-P scanning were not low enough to exclude a diagnosis of PE.

Sensitivity Analysis

Results of sensitivity analysis are shown in Tables 3 and 4. By excluding articles published before 1995, pooled sensitivity (true-positive ratio) decreased from 86.0% to 84.6% (95% CI: 78.3%, 91.6%), while specificity did not change substantially (93.7%). Summary ROC analysis yielded a similar β coefficient (0.588) and a slightly increased *P* value (.118).

DISCUSSION

In the present study, meta-analysis of the contemporary helical CT and V-P scanning literature was performed. On the basis of summary ROC analysis, helical CT and V-P scanning had similar discriminatory power when the high-probability threshold was used. The data also suggest that helical CT has greater discriminatory power than V-P scanning when the inter-

TABLE 2
Clinical Characteristics of Studies Included in Meta-Analysis

A: Helical CT

Study	Mean Age (y)	Women (%)	Time between Tests (h)	Study Design	Enrollment of Patients	Collimation (mm)	Pitch	Location and Size of PE	Verification Bias	Interpretation Bias	True-Positive Findings	False-Negative Findings	False-Positive Findings	True-Negative Findings	No. of Patients
Remy-Jardin et al (8)	34	29	<24	Prospective	Consecutive	5	1	Segmental, main PA	No	No	18	0	1	23	42
Blum et al (9)	43	60	<12	Prospective	NA	5	1	Segmental, main PA	No	No	7	0	0	3	10
Goodman et al (10)	53	40	<24	Prospective	NA	5	1	Subsegmental, main PA	No	No	7	4	1	8	20
Remy-Jardin et al (11)	59	57	<24	Prospective	Consecutive	3–5	1.7	Segmental, main PA	No	No	39	4	0	32	75
Drucker et al (13)	57	53	<12	Prospective	NA	5	1	Segmental, main PA	No	No	8	7	1	31	47
Garg et al (14)	63	4	<24	Prospective	NA	3	2	Segmental, main PA	Likely	No	4	2	0	18	24
Qanadli et al (15)	58	54	<12	Prospective	Consecutive	2–2.5	1	Subsegmental, main PA	No	No	56	3	3	89	151
Ruiz et al (16)	NA	44	<24	Prospective	Consecutive	3	1.6	Segmental-main PA	No	No	21	2	7	31	61
Nilsson et al (17)	53	53	<12	Prospective	NA	3–5	1.3–1.7	Subsegmental, main PA	No	No	30	3	2	55	90

B: V-P Scanning

Study	Mean Age (y)	Women (%)	Time between Tests (h)	Study Design	Enrollment of Patients	Radionuclide Used		Location and Size of PE	Verification Bias	Interpretation Bias	Threshold Used for Data Extraction	True-Positive Findings	False-Negative Findings	False-Positive Findings	True-Negative Findings	No. of Patients
						Ventilation	Perfusion									
PIOPED Investigators (5)	56	55	<24	Prospective	NA	^{133}Xe	^{99m}Tc-MAA	NA	No	NA	1	102	149	14	466	731
											2	207	44	231	249	
											3	246	5	430	50	
Goodman et al (10)	53	8	<48	Prospective	NA	^{99m}Tc-PYP	^{99m}Tc-MAA	Subsegmental, main PA	Likely	No	1	1	10	0	9	20
											2	6	5	4	5	
											3	11	0	9	0	
Garg et al (14)	63	NA	<48	Prospective	NA	^{99m}Tc-DTPA	^{99m}Tc-MAA	Segmental, main PA	Likely	No	1	0	7	0	18	25
											2	7	0	14	4	
											3	7	0	18	0	
Trujillo et al (27)	NA	NA	<48	Prospective	NA	^{99m}Tc-DTPA	^{99m}Tc-MAA	NA	Likely	NA	1	72	100	6	277	455
											2	164	8	177	106	
											3	172	0	280	3	
Woods et al (29)	NA	NA	<48	Retrospective	NA	^{99m}Tc-DTPA	^{99m}Tc-MAA	NA	Likely	No	1	6	7	1	24	38
											2	10	3	13	12	
											3	13	0	23	2	

Note.—NA = not available, PA = pulmonary artery.

mediate-probability threshold was negative. The results of the model comparing helical CT and V-P scanning with normal and/or near normal threshold is confusing because the β coefficient is large, but it borders on being statistically significant ($P = .05$). As will be discussed later, few articles were included in this meta-analysis because of stringent inclusion criteria, which might have decreased our power to detect real difference of two diagnostic tests. To evaluate these issues, it is crucial to take a quantitative view of the data and its interpretation—that is, it is important to interpret the result in terms of effect size rather than in terms of testing (*P* value) (31). Thus, we would rather state that β coefficient of 4.14 is high enough to suggest that helical CT has higher discriminatory power than V-P scanning with use of the normal and/or near-normal threshold.

In two studies (10,14), the diagnostic test performance of both helical CT and V-P scanning were reported in one article, but it was not plausible to directly compare these two modalities for several reasons. In one of these studies, not all patients underwent both tests, and data were insufficient for direct comparison (14). The only study we could use for direct comparison was one in which all patients underwent both helical CT and V-P scanning (10). It might be interesting to directly compare the performance of helical CT and V-P scanning; however, the main aim of the second study was not to compare the performance of the two tests, and there exists a verification bias (eg, patients underwent helical CT on the basis of results of V-P scanning). The direct comparison influenced by verification bias might have caused deviation in the performance of the latter test; thus, it was rather misleading to compare the two tests in this study (23).

TABLE 3
Pooled Sensitivity and Specificity of Helical CT and V-P Scanning with Different Thresholds

Imaging Technique and Year Searched	Threshold	Sensitivity (%)	Specificity (%)
V-P scanning			
1985–2003	1	39.0 (37.3, 40.8)	97.1 (96.0, 98.3)
1985–2003	2	86.0 (83.3, 88.8)	45.5 (43.9, 47.1)
1985–2003	3	98.4 (97.2, 99.5)	4.8 (4.7, 4.9)
Helical CT			
1990–2003	NA	86.0 (80.2, 92.1)	93.7 (91.1, 96.2)
1995–2003	NA	84.6 (78.3, 91.6)	93.7 (91.1, 96.4)

Note.—Data in parentheses are 95% CIs. NA = not applicable.

TABLE 4
Comparison of Summary ROC Analyses between Helical CT and V-P Scanning

V-P Scanning	β Coefficient	*P* Value
Helical CT (1990–2003)		
Threshold 1	0.588 (−1.55, 2.74)	.457
Threshold 2	3.73 (2.56, 4.9)	<.001
Threshold 3	4.14 (−0.002, 8.28)	.05
Helical CT (1995–2003)		
Threshold 1	0.72 (−2.45, 2.9)	.515
Threshold 2	3.63 (2.4, 4.86)	<.001
Threshold 3	3.61 (−9.22, 8.14)	.118

Note.—Data in parentheses are 95% CIs.

Statistical Methods and Results

The summary ROC analysis allows us to compare results of different tests by summarizing sensitivity and specificity results from several studies into a single ROC curve (21). In the past decade, many articles have suggested that helical CT might be more effective than V-P scanning in the diagnosis of PE (32,33), but formal comparison of the two diagnostic tests has yet to be performed. Although previous reports have described helical CT as superior to V-P screening because of its higher sensitivity and specificity, one cannot state that helical CT is more useful for this reason alone. In general, a negative result essentially rules out PE when a test with very high sensitivity is used, and a positive result effectively confirms diagnosis of PE when a test with very high specificity is used (30). In the clinical setting, V-P scanning with a high-probability threshold and high specificity has been used to confirm a diagnosis of PE, and V-P scanning with a normal and/or near normal threshold with high sensitivity has been used to exclude a diagnosis of PE. It is therefore necessary to compare the sensitivity of helical CT with that of V-P scanning by using the normal and/or near-normal threshold to exclude the possibility of PE; it is also necessary to compare specificity to diagnose PE with summary ROC analysis.

There has been criticism of the sensitivities and specificities claimed in earlier helical CT reports (34,35). New diagnostic tests are often described in glowing terms when they are first introduced; however, these tests are often found to be wanting when more experience has been gained (36). This frequently results from limitations in the methods used to evaluate test characteristics. An example of this phenomenon is use of carcinoembryonic antigen (30). Carcinoembryonic antigen was originally considered a very promising tool in the diagnosis of colon cancer; however, carcinoembryonic antigen level was subsequently found to be increased in a wide variety of instances, including in smokers without cancer. The same seems true for helical CT used in the diagnosis of PE. When compared with earlier studies that showed high sensitivity (90%–100%) and specificity (96%–100%), the test performance was lower (sensitivity, 70%; specificity, 91%) in a study performed in 2001 (35). On the other hand, rapid advances in the CT method used in the diagnosis of PE have been made in the past decade. Thus, we performed sensitivity analysis, and the results of ROC analysis did not change when articles published before 1995 were excluded. Moreover, univariate analysis reveals that publication year is not statistically significant.

Another criticism of high accuracy rates, as pointed out in recently published review articles, is that methodologic problems are common in studies used to evaluate helical CT in the diagnosis of PE (eg, several reports were missing key data regarding the methods used to select patients). It is unclear, however, how these methodologic problems have influenced our results (26,37). The second PIOPED study, which is being funded by the U.S. National Institutes of Health, involves evaluation of the accuracy of helical CT in

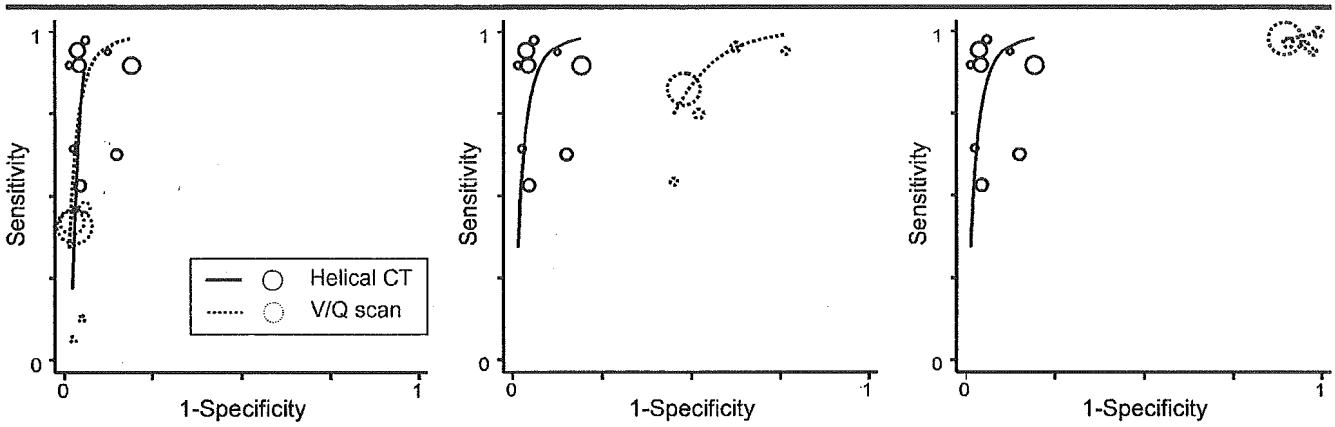

Figure 2. (a) Summary ROC curves for helical CT constructed with a model used to compare V-P scans by using threshold 1. (b) Summary ROC curves for helical CT constructed with a model used to compare V-P scans by using threshold 2. (c) Summary ROC curves for helical CT constructed with a model used to compare V-P scans by using threshold 3. (a–c) Horizontal axis represents the false-positive ratio (1 minus specificity), and the vertical axis represents the true-positive ratio (sensitivity). The size of a plotting symbol is inversely proportional to the variance of an observation. Helical CT has a discriminatory power similar to that of V-P scanning at thresholds 1 and 3, while overall discriminatory power of helical CT is much better than that of V-P scanning at threshold 2.

TABLE 5
Posttest Probabilities of PE with Different Pretest Probabilities and Test Results

Test, Threshold, and Results	Posttest Probability*	Posttest Probability†	Posttest Probability‡
Helical CT			
Positive	.296 (.243, .399)	.841 (.802, .893)	.980 (.974, .987)
Negative	.005 (.003, .006)	.055 (.032, .074)	.347 (.235, .422)
V-P scanning			
Threshold 1			
Positive	.296 (.225, .420)	.841 (.785, .901)	.980 (.971, .988)
Negative	.019 (.018, .020)	.196 (.190, .203)	.522 (.457, .573)
Threshold 2			
Positive	.047 (.044, .049)	.380 (.366, .394)	.848 (.841, .856)
Negative	.009 (.007, .011)	.107 (.085, .129)	.522 (.457, .574)
Threshold 3			
Positive	.031 (.0306, .0313)	.287 (.284, .289)	.786 (.783, .788)
Negative	.010 (.003, .017)	.116 (.036, .185)	.546 (.257, .675)

Note.—Data in parentheses are 95% CIs.
* Pretest probability = .03.
† Pretest probability = .27.
‡ Pretest probability = .78.

the diagnosis of PE in more than 1000 patients and should bring us closer to the solution of these methodologic problems (38).

Clinical Implications and Results

Our work shows that when the high probability threshold is used, helical CT and V-P scanning have comparable discriminatory power; however, when the normal and/or near-normal threshold is used, helical CT had greater discriminatory power than V-P scanning. It is now more important to consider other aspects of the tests. First, the main problem with V-P scanning is that definitive diagnosis can be obtained in less than 30% of patients tested, and the remaining patients need to undergo further testing (5). The use of helical CT would therefore reduce the number of patients subjected to further diagnostic tests (33). In contrast, one should consider the presence of contraindications before performing helical CT. In one report, about 24% of patients suspected of having PE did not undergo helical CT because of contraindications, such as impaired renal function or allergy to contrast agent (35). This is a substantial proportion of patients and is similar to the proportion of patients with inconclusive results of V-P scanning.

Second, as for V-P scanning, large differences (25%–30%) of interpretation among expert readers have been reported, especially in the classification of low- or intermediate-probability scans. In contrast, helical CT has better inter-observer agreement than does V-P scanning (κ value of 0.85 and 0.61, respectively) (12). Third, there are inconsistent results concerning relative cost-effectiveness, with the controversy continuing (32,34). The advantages or disadvantages of either test are crucial in application to the patients when PE is suspected. It is important to judge the advantages and disadvantages of each test and select the most appropriate procedure for a favorable outcome. On the basis of the results of our analyses, we recommend the following strategies: Confirm PE with moderate to high pretest probability, and use either helical CT or V-P scanning, according to high-probability threshold. Exclude PE with low pretest probability, as helical CT is a better test than V-P scanning. If helical CT is not available, V-P scanning with normal and/or near-normal threshold could be an alternative technique. To exclude PE with moderate or high pretest probability or to confirm PE with low pretest probability, avoid V-P scanning by using low-probability threshold.

Limitations

Our review has several limitations. First, because of the nature of meta-analysis, the result is subject to publication bias. Only published reports were exam-

ined, and studies with poor results are less likely to be written, submitted, and accepted. Our results may therefore be biased toward the favorable direction. This tendency should affect helical CT and V-P scanning equally and should not alter our qualitative conclusion. Second, as in all meta-analyses of diagnostic testing, verification bias could be present, since about half of the studies included did not control or mention verification bias. Verification bias occurs when the result of the test influences the decision as to which patients receive the verification test. This can have dramatic results on the sensitivity and specificity of a test (39). We were unable to correct for this bias because the original studies did not provide the necessary information on the entire population tested; in our study, however, covariate analysis did not show a significant difference between studies that did and those that did not control for verification bias. Third, the large degree of variation between observers in reading V-P scans could limit the interpretation of our results when combining studies; this possibility was rejected with a test for homogeneity.

Numerous studies were excluded from analysis, and a smaller number of studies was finally included in meta-analysis compared with previous meta-analysis of helical CT (19). One reason is that some studies used combination reference standards to compare helical CT and V-P scanning (eg, normal results at V-P lung scanning were accepted as an alternative reference standard for the absence of PE) (12,33). The small number of studies included in the current analysis might have decreased our power to detect the true difference. This could not have been avoided, however, because it is usually assumed that the test is being compared with a sole reference standard when meta-analysis of diagnostic tests is conducted. We therefore chose the optimal strategy that pulmonary angiography should be the sole reference standard. For the same reason as mentioned previously, only one multi–detector row helical CT report was included in our analysis. We might have underestimated the test performance of helical CT because the recent introduction of multi–detector row helical CT is expected to offer a further increase in performance, particularly in the ability of physicians to scan larger anatomic volumes with high spatial resolution (40). The stringent inclusion criteria used in the current study, however, should have ensured the quality of our results.

Independently pooled estimates of sensitivity and specificity could be calculated easily; however, these frequently used methods have come under strong criticism because they do not take into account the fact that different studies may have used test thresholds (41). In spite of this, the reason why we pooled estimates of sensitivity and specificity separately is that the results of β coefficients are not always easy for readers to intuitively understand. In a real-world setting, reports of summary ROC analysis present the results of pooled sensitivities; specificities are presented in reports of summary ROC analysis (42,43). Another reason we pooled estimates of sensitivity and specificity separately is that authorities in the field of decision sciences recommend to pool sensitivities and specificities of diagnostic tests and use these data for cost-effectiveness analysis (44). These pooled estimates, however, should carefully be interpreted when used to compare these two diagnostic tests directly.

In conclusion, helical CT has greater discriminatory power than V-P scanning with the normal and/or near-normal threshold in the exclusion of PE, while helical CT and V-P scanning with high-probability threshold had similar discriminatory power in the diagnosis of PE.

References

1. Wells PS, Ginsberg JS, Anderson DR, et al. Use of a clinical model for safe management of patients with suspected pulmonary embolism. Ann Intern Med 1998; 129:997–1005.
2. Hyers TM. Venous thromboembolism. Am J Respir Crit Care Med 1999; 159:1–14.
3. Mills SR, Jackson DC, Older RA, Heaston DK, Moore AV. The incidence, etiologies, and avoidance of complications of pulmonary angiography in a large series. Radiology 1980; 136:295–299.
4. Stein PD, Athanasoulis C, Alavi A, et al. Complications and validity of pulmonary angiography in acute pulmonary embolism. Circulation 1992; 85:462–468.
5. The PIOPED Investigators. Value of the ventilation/perfusion scan in acute pulmonary embolism: results of the prospective investigation of pulmonary embolism diagnosis (PIOPED). JAMA 1990; 263:2753–2759.
6. Stein PD, Hull RD, Raskob GE. Withholding treatment in patients with acute pulmonary embolism who have a high risk of bleeding and negative serial noninvasive leg tests. Am J Med 2000; 109:301–306.
7. Righini M, Goehring C, Bounameaux H, Perrier A. Effects of age on the performance of common diagnostic tests for pulmonary embolism. Am J Med 2000; 109:357–361.
8. Remy-Jardin M, Remy J, Wattinne L, Giraud F. Central pulmonary thromboembolism: diagnosis with spiral volumetric CT with the single-breath-hold technique—comparison with pulmonary angiography. Radiology 1992; 185:381–387.
9. Blum AG, Delfau F, Grignon B, et al. Spiral-computed tomography versus pulmonary angiography in the diagnosis of acute massive pulmonary embolism. Am J Cardiol 1994; 74:96–98.
10. Goodman LR, Curtin JJ, Mewissen MW, et al. Detection of pulmonary embolism in patients with unresolved clinical and scintigraphic diagnosis: helical CT versus angiography. AJR Am J Roentgenol 1995; 164:1369–1374.
11. Remy-Jardin M, Remy J, Deschildre F, et al. Diagnosis of pulmonary embolism with spiral CT: comparison with pulmonary angiography and scintigraphy. Radiology 1996; 200:699–706.
12. Mayo JR, Remy-Jardin M, Muller NL, et al. Pulmonary embolism: prospective comparison of spiral CT with ventilation-perfusion scintigraphy. Radiology 1997; 205: 447–452.
13. Drucker EA, Rivitz SM, Shepard JA, et al. Acute pulmonary embolism: assessment of helical CT for diagnosis. Radiology 1998; 209:235–241.
14. Garg K, Welsh CH, Feyerabend AJ, et al. Pulmonary embolism: diagnosis with spiral CT and ventilation-perfusion scanning—correlation with pulmonary angiographic results or clinical outcome. Radiology 1998; 208:201–208.
15. Qanadli SD, Hajjam ME, Mesurolle B, et al. Pulmonary embolism detection: prospective evaluation of dual-section helical CT versus selective pulmonary arteriography in 157 patients. Radiology 2000; 217:447–455.
16. Ruiz Y, Caballero P, Caniego JL, et al. Prospective comparison of helical CT with angiography in pulmonary embolism: global and selective vascular territory analysis—interobserver agreement. Eur Radiol 2003; 13:823–829.
17. Nilsson T, Soderberg M, Lundqvist G, et al. A comparison of spiral computed tomography and latex agglutination D-dimer assay in acute pulmonary embolism using pulmonary arteriography as gold standard. Scand Cardiovasc J 2002; 36: 373–377.
18. van Beek EJ, Brouwers EM, Song B, Bongaerts AH, Oudkerk M. Lung scintigraphy and helical computed tomography for the diagnosis of pulmonary embolism: a meta-analysis. Clin Appl Thromb Hemost 2001; 7:87–92.
19. Safriel Y, Zinn H. CT pulmonary angiography in the detection of pulmonary emboli: a meta-analysis of sensitivities and specificities. Clin Imaging 2002; 26:101–105.
20. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. Med Decis Making 1993; 13:313–321.
21. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med 1993; 12:1293–1316.
22. Irwig L, Tosteson AN, Gatsonis C, et al. Guidelines for meta-analyses evaluating

diagnostic tests. Ann Intern Med 1994; 120:667–676.
23. Fleischmann KE, Hunink MG, Kuntz KM, Douglas PS. Exercise echocardiography or exercise SPECT imaging? a meta-analysis of diagnostic test performance. JAMA 1998; 280:913–920.
24. Owens DK, Holodniy M, McDonald TW, Scott J, Sonnad S. A meta-analytic evaluation of the polymerase chain reaction for the diagnosis of HIV infection in infants. JAMA 1996; 275:1342–1348.
25. Harvey RT, Gefter WB, Hrung JM, Langlotz CP. Accuracy of CT angiography versus pulmonary angiography in the diagnosis of acute pulmonary embolism: evaluation of the literature with summary ROC curve analysis. Acad Radiol 2000; 7:786–797.
26. Rathbun SW, Raskob GE, Whitsett TL. Sensitivity and specificity of helical computed tomography in the diagnosis of pulmonary embolism: a systematic review. Ann Intern Med 2000; 132:227–232.
27. Trujillo NP, Pratt JP, Talusani S, Quaife RA, Kumpe D, Lear JL. DTPA aerosol in ventilation/perfusion scintigraphy for diagnosing pulmonary embolism. J Nucl Med 1997; 38:1781–1783.
28. Midgette AS, Stukel TA, Littenberg B. A meta-analytic method for summarizing diagnostic test performances: receiver-operating-characteristic-summary point estimates. Med Decis Making 1993; 13:253–257.
29. Woods ER, Iles S, Jackson S. Comparison of scintigraphic diagnostic criteria in suspected pulmonary embolism. Can Assoc Radiol J 1989; 40:194–197.
30. Fletcher RH, Fletcher SW, Wagner EH. Diagnosis. In: Satterfield TS, Taylor C, eds. Clinical epidemiology. 3rd ed. Philadelphia, Pa: Lippincott Williams & Wilkins, 1996; 43–66.
31. Rothman KJ. Random error and role of statistics. In: Epidemiology. An introduction. New York, NY: Oxford University Press, 2002; 113–129.
32. van Erkel AR, van Rossum AB, Bloem JL, Kievit J, Pattynama PM. Spiral CT angiography for suspected pulmonary embolism: a cost-effectiveness analysis. Radiology 1996; 201:29–36.
33. van Rossum AB, Pattynama PM, Mallens WM, Hermans J, Heijerman HG. Can helical CT replace scintigraphy in the diagnostic process in suspected pulmonary embolism? a retrolective-prolective cohort study focusing on total diagnostic yield. Eur Radiol 1998; 8:90–96.
34. Paterson DI, Schwartzman K. Strategies incorporating spiral CT for the diagnosis of acute pulmonary embolism: a cost-effectiveness analysis. Chest 2001; 119:1791–1800.
35. Perrier A, Howarth N, Didier D, et al. Performance of helical computed tomography in unselected outpatients with suspected pulmonary embolism. Ann Intern Med 2001; 135:88–97.
36. Rozanski A, Diamond GA, Berman D, Forrester JS, Morris D, Swan HJ. The declining specificity of exercise radionuclide ventriculography. N Engl J Med 1983; 309:518–522.
37. Mullins MD, Becker DM, Hagspiel KD, Philbrick JT. The role of spiral volumetric computed tomography in the diagnosis of pulmonary embolism. Arch Intern Med 2000; 160:293–298.
38. Kearon C. Excluding pulmonary embolism with helical (spiral) computed tomography: evidence is catching up with enthusiasm (commentary). CMAJ 2003; 168:1430–1431.
39. Roger VL, Pellikka PA, Bell MR, Chow CW, Bailey KR, Seward JB. Sex and test verification bias: impact on the diagnostic value of exercise echocardiography. Circulation 1997; 95:405–410.
40. Ghaye B, Szapiro D, Mastora I, et al. Peripheral pulmonary arteries: how far in the lung does multi-detector row spiral CT allow analysis? Radiology 2001; 219: 629–636.
41. Irwig L, Macaskill P, Glasziou P, Fahey M. Meta-analytic methods for diagnostic test accuracy. J Clin Epidemiol 1995; 48:119–130.
42. Oei EH, Nikken JJ, Verstijnen AC, Ginai AZ, Myriam Hunink MG. MR imaging of the menisci and cruciate ligaments: a systematic review. Radiology 2003; 226:837–848.
43. Visser K, Hunink MG. Peripheral arterial disease: gadolinium-enhanced MR angiography versus color-guided duplex US—a meta-analysis. Radiology 2000; 216:67–77.
44. Hunink M, Glasziou P, Siegel J, et al. Finding and summarizing the evidence: decision making in health and medicine—integrating evidence and values. Cambridge, England: Cambridge University Press, 2001; 214–244.

Efficacy of Lifestyle Education to Prevent Type 2 Diabetes

A meta-analysis of randomized controlled trials

KAZUE YAMAOKA, PHD
TOSHIRO TANGO, PHD

OBJECTIVE — To evaluate the efficacy of lifestyle education for preventing type 2 diabetes in individuals at high risk by meta-analysis of randomized controlled trials, as assessed by incidence and a reduced level of plasma glucose 2 h after a 75-g oral glucose load (2-h plasma glucose).

RESEARCH DESIGN AND METHODS — Through an electronic search, 123 studies were identified. A literature search identified eight studies that met strict inclusion criterion of meta-analysis for 2-h plasma glucose and five studies for the incidence of diabetes. All were randomized controlled trials of ≥6 months with lifestyle education that included a dietary intervention. Subjects were adults diagnosed as being at high risk for type 2 diabetes. The difference in mean reduction of 2-h plasma glucose from baseline to the 1-year follow-up and relative risk (RR) of the incidence of diabetes in the lifestyle education group versus the control group were assessed. Overall estimates were calculated using a random-effects model. Those estimates were confirmed by several models, and the possibility of selection bias was examined using a funnel plot.

RESULTS — Lifestyle education intervention reduced 2-h plasma glucose by 0.84 mmol/l (95% CI 0.39–1.29) compared with the control group. The 1-year incidence of diabetes was reduced by ~50% (RR 0.55, 95% CI 0.44–0.69) compared with the control group. Results were stable and little changed if data were analyzed by subgroups or other statistical models. Funnel plots revealed no selection bias.

CONCLUSIONS — Lifestyle education was effective for reducing both 2-h plasma glucose and RR in high-risk individuals and may be a useful tool in preventing diabetes.

Diabetes Care 28:2780–2786, 2005

Type 2 diabetes is increasing worldwide largely as a result of increasing obesity and a sedentary lifestyle. Nutritional therapy for diabetic patients was recommended by the American Diabetes Association (1). Considering the severity of the illness and low quality of life among diabetic patients, primary prevention for the development of type 2 diabetes is important. For this purpose, lifestyle education (combined diet and exercise) can be considered a powerful tool. Beginning with the impressive study in Da Qing, China (2), the benefits of lifestyle modification have been assessed. Some recent studies based on randomized controlled trials for high-risk subjects revealed the potential for prevention of type 2 diabetes. In a previous study, we conducted a randomized controlled trial of a new dietary education program to reduce plasma glucose levels in Japanese male workers, and we showed that the new dietary education could reduce glucose levels by effecting changes in the total energy intake of individuals at high risk for type 2 diabetes (3). Most current lifestyle education interventions are based on a combination of dietary education with exercise. However, the effects are still controversial. The aim of the present study was to evaluate the efficacy of lifestyle education for preventing type 2 diabetes in individuals at high risk, using a meta-analysis of randomized controlled trials.

From the Department of Technology Assessment and Biostatistics, National Institute of Public Health, Wako, Saitama, Japan.

Address correspondence and reprint requests to Kazue Yamaoka, PhD, Department of Technology Assessment and Biostatistics, National Institute of Public Health, 2-3-6 Minami, Wako, Saitama 351-0197, Japan. E-mail: yamaoka@niph.go.jp.

Received for publication 6 May 2005 and accepted in revised form 8 August 2005.

Abbreviations: DPPRG, Diabetes Prevention Program Research Group; IFG, impaired fasting glucose; IGT, impaired glucose tolerance.

A table elsewhere in this issue shows conventional and Système International (SI) units and conversion factors for many substances.

RESEARCH DESIGN AND METHODS

Study selection and data extraction

The study question was whether a lifestyle education program compared with conventional education improved the overall glucose level or incidence of diabetes in individuals at high risk for type 2 diabetes. Examples of "conventional education" would be usual exercise with or without general information about diet or general dietary advice about healthy food choices on entering the trial.

Outcome measures. To reduce the risk of development of type 2 diabetes, reduction of blood glucose level is necessary. Therefore, the present study considered two outcome measures: the glucose level and incidence of type 2 diabetes. As to the glucose level, the difference in the plasma glucose value 2 h after a 75-g oral glucose load (2-h plasma glucose) between baseline and ≥6 months (mainly 1 year) later was used as an outcome measure. The difference in means of those measures from baseline to 1 year between the lifestyle education intervention and control groups were the effect size of this study. Relative risk (RR), or hazard ratio, for incidence of type 2 diabetes in the lifestyle education intervention group over the control group was another effect size.

Types of participants. Subjects were adults who were diagnosed to be at high risk for type 2 diabetes: impaired glucose tolerance (IGT) (4), impaired fasting glucose (IFG) (5), and borderline (6). The definition of borderline was according to the Japan Diabetes Society (JDS) as follows: normal: fasting plasma glucose <6.1 mmol/l, 2-h plasma glucose <7.8 mmol/l, and 1-h plasma glucose <10 mmol/l; diabetes: fasting plasma glucose ≥7.0 and/or 2-h plasma glucose ≥11.1 mmol/l; and borderline: all remaining values between normal and diabetes. The

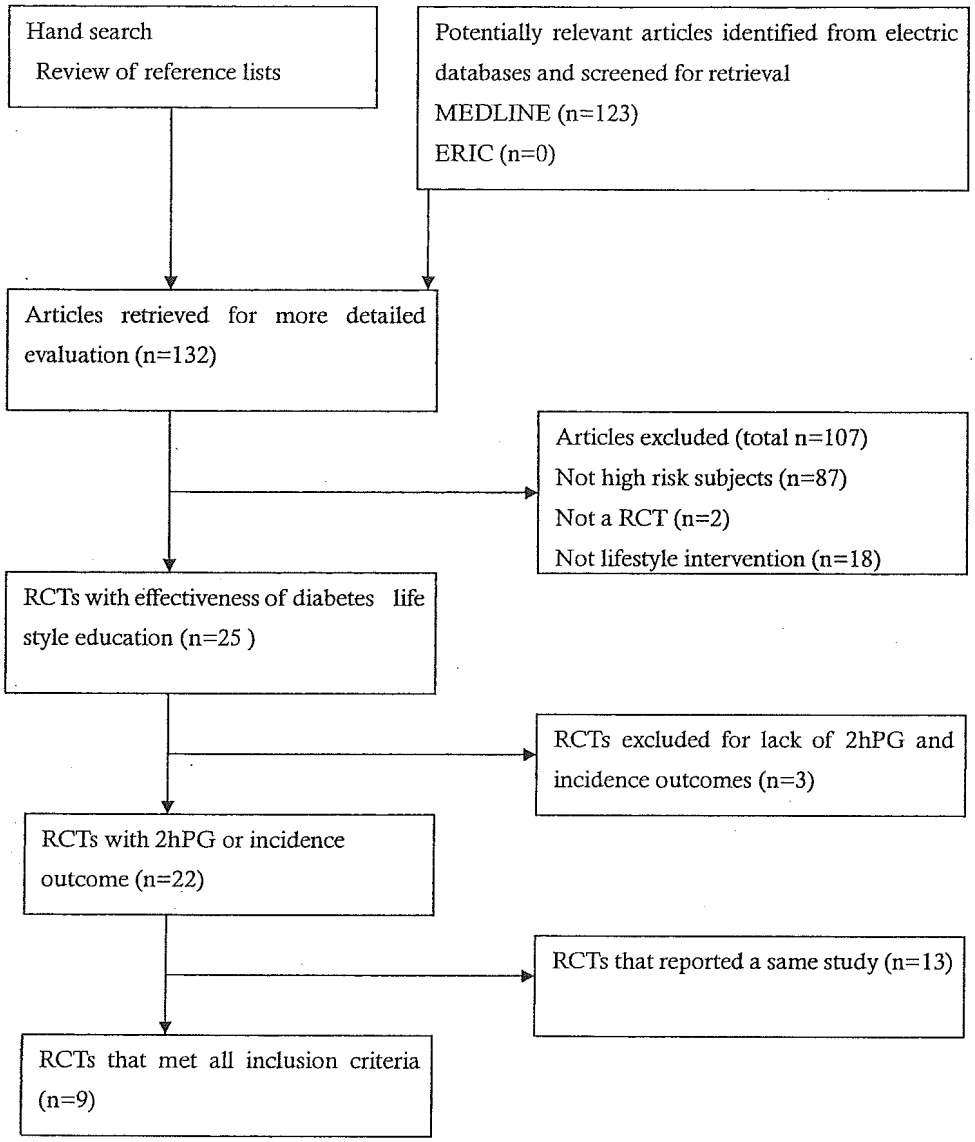

Figure 1—*Systematic review flow diagram.* n = *number of articles. ERIC, Educational Resources Information Center database; 2hPG, 2-h plasma glucose; RCT, randomized controlled trial.*

borderline type corresponds to the sum of IFG plus IGT (6).

Types of studies. Randomized controlled trials that followed patients for ≥6 months were included. Randomization of individuals or clusters of individuals was accepted.

Types of intervention. Lifestyle (combined diet and exercise) or solely dietary education interventions were selected. Control interventions were those described above.

Search strategy for identification of studies. Medline and ERIC (Educational Resources Information Center) databases (January 1966 to November 2004) were searched to identify relevant literature (restricted to the English language). Search terms were free text terms, MeSH (Medical subject heading), and Medline medical index terms. For instance, diabetes, IGT, IFG, borderline, etc. for type 2 diabetes and related conditions; exercise, physical fitness, nutrition, diet, etc., for lifestyle interventions; and prevention and randomized controlled trials were used as search terms.

Statistical analysis

Overall estimates were examined using a fixed-effects model (general variance-based method), a random-effects model (DerSimonian-Laird method) (7), and a Bayesian model with noninformative priors (Monte Carlo Markov chain) (8). A χ^2 test was used to assess heterogeneity among trials. Considering that the fixed-effects model is useful only under conditions of homogeneity and that the power of statistical tests of heterogeneity is low, we planned to use the random-effects model as the primary method irrespective of the test result of heterogeneity. We used the other models for sensitivity analyses. S-plus (9) was used for estimation of the random-effects model and the fixed-effects model, and WinBUGS (10) was used for the Bayesian model (burn-in sample = 1,000, number of Gibbs sampling = 10,000).

The measure of effect size for 2-h plasma glucose is given by the difference between the lifestyle education intervention group and control group (Δ) for each individual study, which is equal to $\Delta_i - \Delta_c$, where Δ_i and Δ_c are mean differences from baseline to end point (basically at 1 year) in 2-h plasma glucose between, respectively, the lifestyle education intervention and control groups. When the SD of the difference from baseline to end point was not given in the literature, it was calculated using SD_{pre} (SD of the baseline 2-h plasma glucose) and SD_{post} (SD of the end point) for each group, using the formula $SD^2 = SD^2_{pre} + SD^2_{post} -$

Table 1—*Characteristics of the nine randomized controlled trials**

Study (ref. no.)	Randomized subjects	Inclusion criteria	Follow-up duration (years)	Diabetes incidence (*r*/*n*)†		2hPG (mmol/l)	
				Control	Intervention	Baseline (means ± SD)	Difference from baseline at 1 year (means ± SD) (*n*)‡
Pan et al. (2)	577§	M&F, IGT	6	90/133	58/130	C: 9.03 ± 0.89 L: 9.11 ± 0.93	C: 3.96 ± 3.82 (133)‖ L: 1.65 ± 3.16 (130)
Wein et al. (13)	200	Female, IGT	4.25	7/100	6/100	C: 9.8 ± 0.74‖ L: 9.9 ± 0.74	C: 0.1 ± 1.94 (96)‖ L: −0.1 ± 2.19 (97)
Lindahl et al. (14)	186	M&F, BMI >27, age 30–60, IGT	1	NA	NA	C: 8.0 ± 11.09¶ L: 7.5 ± 6.99	C: −0.30 ± 2.75 (93)¶ L: −0.68 ± 1.95 (93)
Oldroyd et al. (15)	78	M&F, age 24–75, IGT	0.5	NA	NA	C: 9.2 ± 0.9 L: 9.1 ± 0.9	C: −0.5 ± 1.8 (32)# L: −0.7 ± 1.9 (35)
Tuomilehto et al. (16)	522	M&F, BMI >25, age 40–64, IGT	6**	51/257	22/265	C: 8.9 ± 1.5 L: 8.9 ± 1.5	C: −0.3 ± 2.2 (250) L: −0.9 ± 1.9 (256)
Swinburn et al. (17)	176	M&F, age ≥40, IGT + (2hPG 7.0–7.8 mmol/l)	1	NA	NA	C: 7.5 ± 2.4 D: 7.9 ± 2.5	C: 0.74 ± 2.76 (70) D: 0.01 ± 2.68 (66)
Mensink et al. (18)	114	M&F, BMI >25, age ≥40, IGT	3	NA	NA	C: 8.6 ± 1.48 D: 8.8 ± 2.06	C: 0.2 ± 2.23 (55) L: −0.8 ± 2.06 (47)
Watanabe et al. (3)	173	Male, age 35–70, borderline	1	6/87	3/86	C: 7.3 ± 1.7 D: 8.2 ± 1.5	C: 0.67 ± 1.74 (77) D: −0.76 ± 1.36 (79)
DPPRG (19)	3,234§	M&F, age ≥25, BMI >24 (Asian >22), 27 centers, IFG	2.8	313/1,082‡	155/1,079‡	C: 9.1 ± 0.9 L: 9.1 ± 0.9	NA

*Nine studies were reported in 22 published articles. One article is listed as a representative of the relevant study. †Incidence of type 2 diabetes: *r*/*n* = (number of cases divided by total number of analyzed subjects). ‡Except for the studies by Pan (2) = 6 years, Wein (13) = average 4.25 years, and Oldroyd (15) = 6 months. §Including other intervention types. ‖SD for the mean difference was calculated using SDs in each point. ¶SD was calculated using 95% CIs. #Difference from baseline at 6 months. **Strong intervention was performed during the 1st year. ††Calculated from incidence. C, control; D, solely dietary education intervention; L, lifestyle education (combined diet and exercise) intervention; M&F, male and female.

$2rSD_{pre}SD_{post}$, where r is the correlation between the baseline and end point groups. Because no study reported r, and its true value is unknown, we consulted our past study data and used $r = 0.5$. For this, a sensitivity analysis was performed, using $r = 0.3$ and $r = 0.7$. If the 95% CI was shown instead of SD, SD was calculated using the formula $SD = (\sqrt{n})$ (95% CI_{upper} − 95% CI_{lower}) ÷ 4, where n denotes sample size of a group.

Net change in 2-h plasma glucose or RR is shown for each individual study, with lines extending from circles representing 95% CIs in the Forest plot. A cumulative meta-analysis by the random-effects model (11) was also performed to determine at which point (when sufficient evidence was available) to demonstrate a beneficial lifestyle education intervention effect. Subgroup analysis by intervention type, i.e., diet versus lifestyle (combined diet and exercise) and follow-up duration (<1 vs. ≥2 years), was conducted as a sensitivity analysis. The selection bias was visually examined using the funnel plot.

RESULTS — Following the QUOROM guidelines (12), Fig. 1 depicts the flow diagram for this review. Eight studies (2,3,13–18) met strict inclusion criteria for the analysis of 2-h plasma glucose and five studies (2,3,13,16,19) for the analysis

Table 1—*Continued*

Age (years)	BMI (kg/m^2)	Type of intervention	Dietary education	Exercise education	Control
45	26	Dietary + exercise	Reducing energy intake	Increase leisure physical exercise at least 1 unit/day	General instructions for diet and/or increased leisure physical activities
39	25	Dietary + exercise	Standard diet advice sheet with telephone contact (three per month)	Emphasizing need for regular exercise	Regular exercise and standard diet advice
55	31	Dietary + exercise	Low-fat, high-fiber diet	Regular exercise with a program implemented during a 1-month stay at a wellness center that included intense dietary learning sessions	Standard program including counseling session for 30–60 min conducted by a specially trained nurse
58	30	Dietary + exercise	Regular diet counseling from a dietician	Physical activity counseling from a physiotherapist	General instructions for diet and/or increased leisure physical activities
55	31	Dietary + exercise	Individualized dietary counseling from a nutritionist	Circuit-type resistance training sessions and advice on increasing overall physical activity	General dietary and exercise advice at baseline and an annual physician's examination
52	29	Dietary alone	Reduced-fat diet and participation in monthly small-group education session for 1 year	None	General dietary advice about health food choices
57	29	Dietary + exercise	Regular dietary advice	Stimulated to lose weight and increase physical activity with visits scheduled at regular intervals	Brief information about the beneficial effects of a healthy diet and increased physical activity
55	24	Dietary alone	Reducing energy intake, especially at dinner	None	Conventional group counseling
50	34	Dietary + exercise	Weight reduction through a healthy low-calorie, low-fat diet	Engage in physical activity of moderate intensity by individualized curriculum by case managers	Written information for standard lifestyle recommendation and an annual 20- to 30-min individual session emphasizing importance of a healthy lifestyle

of RR. In this process, one study was selected among studies that published results from the same trial, and one intervention (having priority on lifestyle intervention over the diet-alone intervention) was selected from a study. The general characteristics and outcomes of the studies are shown in Table 1.

Type of intervention

Lifestyle education interventions of the selected studies varied widely. Lifestyle education (combined diet and exercise) was conducted in seven studies (2,13–16,18,19), and a solely lifestyle education intervention was carried out in two studies (3,17). Details of the type of intervention are summarized in Table 1.

2-h plasma glucose

Two studies (2,13) did not report the SD of the differences from baseline to end point, so the SD was calculated from SD_{pre} and SD_{post}. Two studies (13,14) showed the 95% CI instead of SD; thus, SD was calculated from 95% CI_{upper} and 95% CI_{lower}. In the eight studies (2,3,13–18) in which the 2-h plasma glucose level was determined, evidence of heterogeneity among the studies was shown ($P < 0.001$). Figure 2 shows the net change in 2-h plasma glucose, results of cumulative meta-analysis, and overall estimates for 2-h plasma glucose by several models. The estimates from the random-effects model are shown, with lines extending from quadrangular symbols representing 95% CIs. The ranges of 95% CIs of the overall estimates for several models are shown with the solid line between the diamond symbols in the figure. In calculating overall estimates for 2-h plasma glucose, the results were insensitive to r in the range we expected (0.3–0.7); therefore, data are presented with the value of $r = 0.5$. Cumulative analysis indicated that from the last four studies, overall estimates became significant. Overall, a

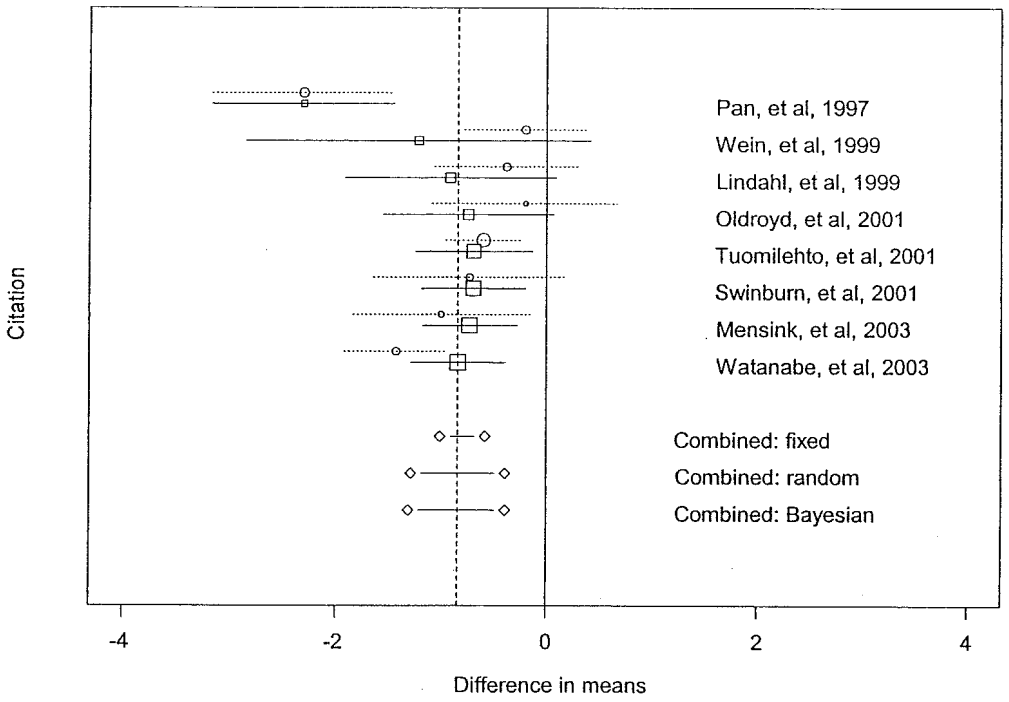

Figure 2—*Forest plot for the net change in 2-h plasma glucose in eight randomized controlled trials of the effects of lifestyle education, with their 95% CIs (individual and cumulative meta-analysis). Net change in 2-h plasma glucose is shown for each individual study, with dotted lines extending from circles representing 95% CIs. Cumulative meta-analysis by the random-effects model in 2-h plasma glucose is shown by each individual study (sequentially cumulated), with solid lines extending from quadrangles representing 95% CIs. The ranges of 95% CIs of the overall estimates are shown for several models with solid lines between the diamonds.*

1-year lifestyle education intervention reduced 2-h plasma glucose by 0.84 mmol/l (95% CI 0.39–1.29) compared with the control intervention, as determined by the random-effects model. Concordant results were obtained by other models, i.e., a 0.80 mmol/l (0.58–1.01) reduction was estimated by the fixed-effects model and a 0.84 mmol/l (0.39–1.32) reduction by the Bayesian model. All of the overall estimates denoted a significant reduction of 2-h plasma glucose in the lifestyle education intervention groups compared with control groups.

Because there was evidence of heterogeneity in this combined analysis, subgroup analyses were conducted to analyze sensitivity. Overall estimates of 2-h plasma glucose were obtained according to the length of the study (1 year for five studies and >1 year [6 and 4.25 years] for two studies) and by the types of intervention (lifestyle education for six studies and solely dietary education for two studies.) Excluding studies that exceeded 1 year (two studies), the results still showed a significant reduction in 2-h plasma glucose, except those for the Bayesian model.

A funnel plot of sample size against the effect size was examined (figure not shown). From observations of data, selection bias did not largely affect the results of the present study. In addition, the related factors of mean age, study publication year, baseline value of 2-h plasma glucose, and BMI varied, and these factors were visually examined. From an observational point of view, the results detected no bias (figures not shown).

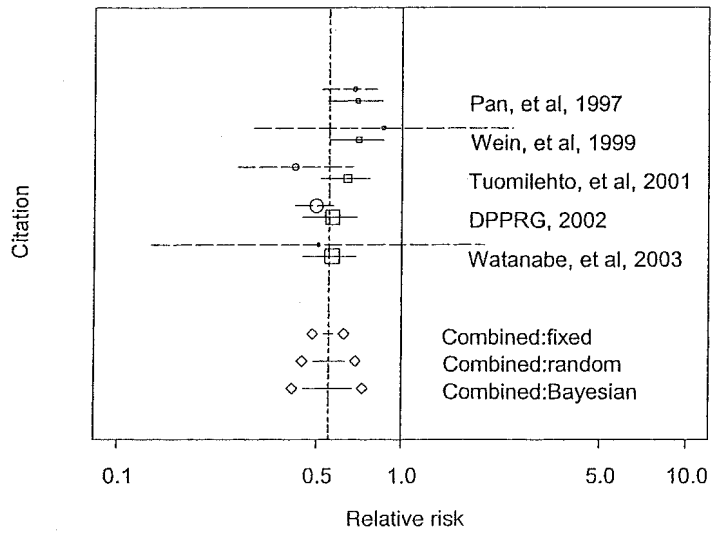

Figure 3—*Forest plot for RR in five randomized controlled trials of the effects of lifestyle education, with their 95% CIs (individual and cumulative meta-analysis). For explanation of the figure, see the legend to Fig. 2.*

RR

In the five studies (2,3,13,16,19) in which the incidence was obtained, analysis showed no evidence of heterogeneity among studies (P = 0.145). Figure 3 shows RRs of each study, the result of cumulative meta-analysis, and overall RRs by several models. The cumulative meta-analysis indicated significant effects in all cases. All of the results indicated that lifestyle education groups had a relatively lower incidence than control groups. The risk of incidence of type 2 diabetes in the lifestyle education intervention group was reduced by ~50% (RR = 0.55 [95% CI 0.44–0.69]) compared with the control intervention group by the random-effects model. The results from other models were similar. Specifically, RR was estimated as 0.55 (0.48–0.63) by the fixed-effects model and 0.55 (0.41–0.74) by the Bayesian model. Because there was one mega-study conducted by the Diabetes Prevention Program Research Group

(DPPRG) (19), we excluded it from analysis. Although the number of subjects was small except for the mega-study, meta-analysis of small trials was shown to be concordant with the results of the mega-study when we examined the fixed-effects model, the random-effects model, and Bayesian model.

CONCLUSIONS — This meta-analysis provided evidence of the efficacy of lifestyle education for individuals at high-risk of type 2 diabetes in reducing 2-h plasma glucose and RR. It reduced 2-h plasma glucose by ~0.84 mmol/l (95% CI 0.39–1.29) and also the incidence of type 2 diabetes by ~50% (RR 0.55 [0.44–0.69]) compared with the control group, as determined by the random-effects model. Significant effects were also obtained by other models. Although the interventions and methods of lifestyle education varied in these studies, these results indicate that lifestyle education as well as a solely dietary education improved 2-h plasma glucose and reduced the risk of type 2 diabetes in high-risk individuals.

Although lifestyle education for high-risk subjects is an accepted cornerstone of prevention of type 2 diabetes as well as treatment of type 2 diabetes, a formal and systematic overview of its efficacy and method of delivery has not been available. Our study provides evidence of a relationship between lifestyle education in high-risk subjects and the prevention of type 2 diabetes.

Several meta-analyses have been published on the effects of lifestyle education on GHb for diabetic patients (20), low-glycemic index diets in the management of diabetes (21), and glucose and insulin responses to dietary chromium supplements (22). Although the purpose, methods, and types of subjects differed, there was evidence that not only clinical care but also lifestyle education is effective. Our study aimed at examining lifestyle education for those at high risk of type 2 diabetes. Considering the poor quality of life of diabetic patients, preventing the development of this disease is important, and much more attention should be paid to lifestyle education.

Many individuals at high risk for diabetes are designated as having what is now called metabolic syndrome, and, recently, considerable attention has been paid to this syndrome. The primary end points of the randomized controlled trials analyzed in the present study were designed as 2-h plasma glucose and/or incidence of type 2 diabetes. Therefore, we cannot examine the effects of lifestyle education on the metabolic syndrome. Obesity is one component of the metabolic syndrome. Many studies examined BMI as one of the secondary end points. Some of the individual studies (13–15,17) did not find a significant effect of lifestyle education on 2-h plasma glucose, but they did find that it affected BMI. This means that a weak effect of lifestyle intervention on weight loss may exist. Further study of the metabolic syndrome is needed to define effective interventions for this condition.

Many of the studies included in this meta-analysis involve only a small number of subjects, with the exception of one mega-trial (19), which was used for analysis of RR in this report. The results, when excluding the mega-study, were also significant. The findings suggest the clinical benefits of lifestyle education. There has been extensive discussion of the differences between meta-analyses and mega-trials (23). Selective nonpublication of negative trials seems to be a likely explanation for that. Our results suggest that the meta-analyses of small trials is concordant with the results of the DPPRG, for which we examined the random-effects model as well as the fixed-effects model and Bayesian model.

The strengths and limitations of this meta-analysis should be considered. Our study has several strengths. As far as we know, this is the first study to examine the effects of lifestyle education for individuals at high risk of type 2 diabetes by meta-analysis, although the education reported in the studies was not uniform. We also focused attention on two types of viewpoints: 2-h plasma glucose and incidence. Considering that those with higher values for 2-h plasma glucose are more likely to develop diabetes, it is meaningful that both glucose level and incidence indicated the effect of lifestyle education when compared with control subjects.

This study has several important limitations. This analysis was confined to English-language articles, which could introduce bias. Furthermore, only randomized controlled trials were included, which could also introduce bias. However, considering that the quality of studies of lifestyle education as well as solely dietary education may be affected by many confounding biases, these limitations may be acceptable. From the visual observations by plots on the effect of lifestyle education against the factors on the effect size, the results were not greatly affected by those factors. Publication bias is always a concern in meta-analyses. We performed electronic searches, including a hand search, and examined by funnel plot sample size against effect size. The funnel plot suggested little influence from publication bias on the effect size. Although it may be small, we cannot deny the possibility of selection bias. Our study has a limitation in that the follow-up period extended for ≥6 months; however, this may be acceptable because an earlier assessment could be biased as a result of changes made only because subjects were conscious of being studied. In the prevention of diabetes, maintaining long-term control is warranted. Another limitation, which was the variability of lifestyle education, was examined by subgroup analyses. Although the quality of lifestyle education varied, the results indicated that it was effective.

Taking these limitations into account, the meta-analysis provided objective evidence that lifestyle education for reducing 2-h plasma glucose and the incidence of type 2 diabetes in groups of high-risk individuals is effective and may be a useful tool for preventing type 2 diabetes. Approaches that include lifestyle education with the goal of preventing the development of type 2 diabetes should be given more attention.

Acknowledgments — This study was supported by a Ministry of Health, Labor, and Welfare in Japan grant-in-aid for scientific research 2003–2005.

References

1. American Diabetes Association: Nutrition principles and recommendations in diabetes. *Diabetes Care* 27 (Suppl. 1):S36–S46, 2004
2. Pan XR, Li GW, Hu YH, Wang JX, Yang WY, An ZX, Hu ZX, Lin J, Xiao JZ, Cao HB, Liu PA, Jiang XG, Jiang YY, Wang JP, Zheng H, Zhang H, Bennett PH, Howard BV: Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance: the Da Qing IGT and Diabetes Study. *Diabetes Care* 20: 537–544, 1997
3. Watanabe M, Yamaoka K, Yokotsuka M, Tango T: Randomized controlled trial of a new dietary education program to prevent type 2 diabetes in a high-risk group of Japanese male workers. *Diabetes Care* 26: 3209–3214, 2003 [erratum in *Diabetes Care* 27:856, 2004]