Published in final edited form as:

J Am Geriatr Soc. 2015 April ; 63(4): 770–775. doi:10.1111/jgs.13337.

Interactive Voice Response Version of the Late-Life Function and Disability Instrument

Feng-Hang Chang, ScD, OT^{1,2}, Nancy K. Latham, PhD, PT², Robert H. Friedman, MD³, and Alan M. Jette, PhD, MPH, PT²

¹Graduate Institute of Injury Prevention and Control, College of Public Health and Nutrition, Taipei Medical University, Taipei, Taiwan

²Health and Disability Research Institute, Boston University School of Public Health, Boston, MA, USA

³Department of Medicine, Boston University School of Medicine, Boston, MA, USA

Abstract

OBJECTIVES—To develop an interactive voice response (IVR) version of the Late Life Function and Disability Instrument Computer Adaptive Test (LLFDI-CAT) and evaluate its reliability and acceptability among older adults.

DESIGN—The IVR system (IVRS) was embedded within the LLFDI-CAT program. To test the test-retest reliability and concordance of the IVR version of LLFDI-CAT with the telephone interviewer form (TIF), participants received the two versions of LLFDI-test at baseline, and at 1-week follow-up.

SETTING—Community

PARTICIPANTS—Community dwelling adults aged 65 and older (N=50)

MEASUREMENTS—The LLFDI is a patient-reported outcome (PRO) measure developed to assess function and disability in older adults.

RESULTS—The IVR administered version of the LLFDI-CAT showed acceptable overall test-retest reliability (ICC=0.79–0.80) and concordance with the TIF (ICC=0.74–0.97). Although most participants preferred the TIF, the majority did not find the IVR version more difficult to use.

Corresponding author: Feng-Hang Chang, ScD, OT, Graduate Institute of Injury Prevention and Control, Taipei Medical University, 250 Wu-Hsing Street, Taipei City, Taiwan 110, Phone: 886-22736-1661#6573, fhchang@bu.edu.

Conflict of Interest: Dr. Jette holds stock in CREcare, LLC, has a small business that disseminates CAT instruments including the LL-FDI. All other authors have no other conflicts of interest to disclose.

Author Contributions: Dr. Chang took the role of literature review, data analysis, interpretation of data, and manuscript preparation. Dr. Latham was responsible for the study design, data collection, and served as principle investigator for the project. Dr. Friedman contributed his knowledge of the development and application of the IVR technology. Dr. Jette contributed to the study design, provided mentorship, and contributed critical feedback on the manuscript. All authors had full access to all of the data in the study and took responsibility for the integrity of the data and the accuracy of the data analysis.

Sponsor's Role: The sponsors had no role in the design, analysis or interpretation of these data. Funding received from National Institute on Aging (Grant number: 5P30AG031679)

CONCLUSION—The IVR version of LLFDI-CAT achieved reliability levels that were comparable with the TIF version. Future work is needed to improve the IVR design to better fit older adults' needs and preferences.

Keywords

automated telephone system; aging; interactive voice response; disability; participation

INTRODUCTION

The Late-Life Function and Disability Instrument (LLFDI) is a patient-reported outcome measure developed to assess function and disability in community-dwelling older adults.¹ Unlike many other measures that focus on activities of daily living, the LLFDI assesses both a person's inability to perform discrete physical tasks and his/her inability to participate in major life tasks and social roles, based on the Nagi's disablement model² and the World Health Organization International Classification of Functioning, Disability, and Health (ICF).³

Since its development in 2002, LLFDI has been used as an outcome measure in over 70 studies.⁴ The original LLFDI was a fixed-form, which has raised concerns over respondent burden and administration costs.¹ A revised LLFDI, which applies computer adaptive testing (CAT) techniques for its administration, was developed to solve these issues and has demonstrated promising psychometric properties.⁵ A CAT tailors administration of a test to the current ability level of each subject so that only items that are appropriate to an individual are administered, thus minimizing the number of items administered without sacrificing measurement precision.⁵ To further improve the feasibility of administration, we applied the telephone and interactive voice technology (IVR) to the LLFDI-CAT.

IVR is a computer-automated telephone system that uses interactive script, and predetermined call-flow algorithms that control the sequencing of questions.^{6, 7} The respondents use touchtone telephone keypad or verbal responses to interact with the system, and the answers are automatically entered into a database allowing rapid and computer-generated interpretation, and report generation. Data entered into the database is decoded and translated into tables or graphs to be used by clinicians or researchers. IVR systems have been used for management of diverse health conditions.^{6, 8, 9}

Numerous advantages of IVR technology have been described in the literature, including: 1) Cost-effective administration; 2) 24 hour, 7 days a week access for interviewing; 3) no interviewer bias or administration errors; 4) automatic response validation that reduces responses outside of allowed range or logical inconsistencies; 5) immediate availability and automated dissemination of results; 6) efficient screening of large numbers of people; 7) time-saving, automatic redialing of unanswered calls until recipient is reached.^{10–14} To our knowledge, it has never been applied in measuring late-life disability and little evidence comparing the psychometric properties of IVR and telephone interviewer administered assessments, even though several studies have compared IVR favorably to paper-and-pencil and clinician face-to-face interviews.^{10, 11, 14}

The aims of this study included: (1) develop an IVR system for administering the LLFDI-CAT; (2) determine the test-retest reliability and concordance of IVR to a telephone-interviewer administered instrument; (3) compare the acceptability of using the IVR with a telephone interviewer-administrated approach in older adults.

METHODS

Measure

The LLFDI contains two summary scales: a Function domain scale and a Disability domain scale. The Function domain scale includes questions asking “how much difficulty do you currently have doing a particular activity?” or “how much help from another person do you currently need doing a particular activity?” (Response options range from 1: none at all to 4: unable to do) The Disability domain scale includes questions asking “Because of your physical or mental health, to what extent do you feel limited in doing a particular activity?” (Response options range from 1: not at all to 4: Completely).^{15, 16}

The original LLFDI-CAT Function scale assesses participants’ difficulty in performing 32 functional tasks. The Disability scale assesses frequency (11 items) and limitation of participating in 16 activities.

The development of the IVR for the LLFDI-CAT includes the following steps: (1) preparing design specifications, (2) writing the full script, (3) computer programming the dialogue content, logic, report layout, and system database, (4) voice recording the dialogues, and (5) debugging the system.

In the first step, the design specifications were written for administration of the LLFDI over a telephone linked computer system,¹⁷ which is a computer-based telecommunications system that converses with people via phone. The telephone linked computer system combines an IVR subsystem for generating speech over the telephone, a speech recognition subsystem for recognizing what the user is saying, a database management subsystem for storing and managing system and user data, and a conversation control subsystem that controls the content and flow of individual conversations with users. The telephone linked computer carries out a series of totally automated telephone conversations with the users.

In the second step, the investigators wrote the telephone linked computer dialogues based on the LLFDI-CAT into English phrases and sentences.

The third step was to implement LLFDI IVR into the telephone linked computer that controls and executes the conversations with the patients, and constructs and maintains the system database. This entailed computer specification of the conversation logic, design of the system’s database, and entering the dialogue content and other data into the system’s database.

The final step was voice recording the dialogue using a fourth generation of Visual VoiceTM, to construct the computer representation of the logic of the conversations. After these steps were completed, the IVR application was fully programmed and the software was thoroughly tested and debugged.

The SF-36 physical functioning scale (PF-10) was used to collect participants' baseline functional status.¹⁸ PF-10 consists of 10 items measuring perceived limitations in a variety of physical activities on a 3-point response scale: from 1 (limited a lot) to 3 (not limited at all). The summary score was linearly transformed to range between 0 and 100, with higher scores indicating higher physical functioning.

Design

To assess the concordance and reliability of the LLFDI-CAT, all participants received the IVR and the telephone interviewer form (TIF) of the LLFDI-CAT. The order of mode of administration was randomly assigned at the baseline assessment. One week later, the participants were retested using both versions of the LLFDI-CAT again. The 1-week timeframe was used since a person's status was likely to remain stable during that time period and memory would not invalidate the retest. To avoid bias in the evaluation of the measures due to the order effect, the order of the administration modes was reversed at the follow-up assessment.

At the end of the follow-up assessment, participants were asked to rate the preference and difficulty of using the two administration modes using Likert rating scales and open-ended questions.

Participants

Fifty-one community-dwelling older adults with physical functional limitations and/or disabilities were recruited by telephone from the greater Boston area using a subject registry from prior studies. Eligibility criteria included: (1) speak English; (2) be at least 65 years old; (3) understand and respond to interview questions; and (4) have two or more limitations on the Physical Function Scale of the SF-36 instrument. Fifty participants completed the study. The mean age of participants was 77.4 years (SD= 8.1 years); 39 (78.4%) were females. Fifty subjects completed all study assessments.

Data Analysis

The interclass correlation coefficient (ICC) was used to estimate test-retest reliability of the IVR and TIF versions of LLFDI-CAT and to address the concordance of scores between the two modes.^{19, 20} Test-retest reliability was estimated using the baseline and 1-week follow-up test scores of each of the administration modes. The concordance between the two administered modes was examined at both baseline and at follow-up. The ICCs were interpreted as poor (0.0–0.50), moderate (0.51–0.75), good (0.76–0.90), or excellent (0.91–1.00).²⁰

Participants' responses regarding the preference and difficulty of using the two administration modes were summarized at the end of the study.

RESULTS

Table 1 displays the characteristics of two groups of the study sample. The average age was 78.5 years in the IVR first group and 76.4 years in the TIF first group. Both groups

contained 76% females, and 100% white. The mean PF-10 score of the two groups were 71.8 and 69.7, respectively.

The test-retest reliability of the IVR and TIF of LLFDI-CAT and the concordance between the two versions were shown in Table 2. Overall, the IVR version showed good test-retest reliability (ICC=0.79 to 0.80). When broken into subgroups by order of mode of administration, the test-retest reliability was moderate in the group with IVR first (ICC=0.54–0.65) but was excellent in the group with TIF first (ICC=0.94–0.97).

The concordance between the two versions of LLFDI-CAT was good at baseline (ICC=0.74–0.83), and was excellent at 1-week follow-up (ICC=0.96–0.97). For the two subgroups with the IVR administered first and with TIF administered first, the concordance was moderate to excellent at baseline (ICC= 0.67–0.97), and excellent at follow-up (ICC=0.95–0.98).

Preference and difficulty of using the two administered methods

Figure 1 demonstrates the participants' preference and perceived difficulty over the two administered methods. All participants have rated their preferences on 5-point Likert scales; 26 (52%) had strong preference for the TIF; 11 (22%) had slight preference for the TIF (see the upper chart in Figure 1). No participant showed preference for the IVR version.

According to the responses to open-ended questions, many participants felt the TIF was more personal and made them feel comfortable to respond. Some of them clearly stated that they dislike "talking to computers" and were longing for human contacts. A number of participants said they liked the TIF because they can explain their answer better and add more comments if necessary. A participant wanted to ask questions of the interviewer during the assessment and found he could not ask for clarification when using the IVR version. Participants did not like that they had to wait for all the answers to be read by the computer before they gave a response.

Thirty four (68%) participants found no difference in difficulty between the two methods; 13 (26%) found IVR was slightly more difficult to use; and 3 (6%) found IVR was much more difficult to use(see the lower chart in Figure 1). Those who found that IVR was more difficult to use described the primary challenge was that they had difficulty to understand the recorded voice, and IVR sometimes could not understand their responses. Additionally, a few participants had hearing problems, and they suggested that the recorded voice should be louder and clearer.

DISCUSSION

This study is the first study that applied IVR technology to assess function and disability in community dwelling older adults and that combined IVR with CAT technology to administer a functional assessment and compared its psychometric properties with a TIF version.

Taken together, the IVR administered version of the LLFDI-CAT showed excellent test-retest reliability (ICC=0.79 to 0.80). The ICCs between IVR and TIF version of LLFDI-

CAT were good at baseline and excellent at follow-up, which indicated high concordance between the two administration modes. These findings provide supportive evidence for applying IVR technology to administer a functional assessment.

However, changing the order of administration modes seems to affect the test-retest reliability: the IVR assessment was less reliable when the IVR was administered first (ICC=0.54–0.65) but excellent when the TIF was administered first (ICC=0.94–0.97). This finding was unexpected and may be explained by the participants' unfamiliarity with the IVR operational mechanism. When the participants were exposed to the IVR first, it is possible that they did not fully understand the operational system which distracted them from focusing on the content of the questions, but were able to perform well in later interviews once they were more familiar with the IVR version. On the other hand, when the TIF version was administered first, participants did not need time to get familiar with the IVR operational system and could focus more on the content of the questions being asked. By the time this group was introduced to the IVR system, they were already familiar with the questions and were therefore more consistent in their responses. Based on our findings, we also suggest that giving consumers some practice of using IVR to help them get familiar with the system before administering a questionnaire. Furthermore, a more humanized and personal design with more lively interaction may enhance older adults' acceptance toward IVR and decrease operational challenges.

The order effect between the IVR and personal interviews has not been widely explored in the literature. One study compared a computer-administered questionnaire and personal interviews in assessing oral-health behaviors.²¹ It found that a few items demonstrated less concordance between the two administration modes when the computer-administered interview came first but not when the personal interview administered first.²¹ The reason for this finding remained unclear. Thus, we suggest that more qualitative investigation may be useful to understand better the cause(s) of the order effect between the IVR and personal interviews.

About half of the participants in this study revealed preference to the personal interviews over the IVR even though the majority of them (68%) did not find the IVR version more difficult to use. This finding is in line with literature and may reflect the widespread dislike of computerized telephone system in daily life.^{22, 23} Many participants in our study mentioned that they preferred to talk to a real person on the phone since it was more personal and engaged. They also felt more comfortable to interact with human beings than talking to a computer and perceived more flexibility to ask questions or interrupt live conversations. Previous studies suggest that older people particularly prefer to talk to real personnel than having an IVR interview.^{24, 25} The attributing factors include older people's unfamiliarity with IVR technology and declined cognitive ability such as the ability to process information and maintain concentration, sensory or motor limitations (i.e., hearing problems).²⁶ All these challenges may affect seniors' acceptability of IVR.

Although the acceptability of IVR in older adults remains a challenge, the benefits of this system should not be overlooked. In the past decades, IVR has been applied in diverse healthcare services, such as health and functional assessment, behavioral intervention, and

patient monitoring, and has demonstrated distinct benefits.¹² The most noticeable advantages of IVR were its low administration cost and convenience. Unlike live interviewers that can only reach out to one consumer at one time, IVR can administer multiple simultaneous calls by the pre-recorded script and save tremendous administration and interviewer-training cost.¹¹ Consumers can receive the call anytime and any location as their convenience. Additionally, IVR uses consistent and standardized process to obtain information to prevent interviewer bias and administration errors.¹¹ With IVR, consumers can also provide sensitive information more comfortable compared to talking to a real person.¹¹ An IVR system could provide a cost-effective way to allow frequent monitoring of consumers during phases when they are at high risk of functional decline (i.e. right after discharge from hospital) or over extended periods of time (i.e. monitoring functional change at regular intervals over several years). The high cost of live interviews would make frequent and/or long duration assessments of function not feasible for many health systems or research studies.

To address the pros and cons of IVR, further investigation is needed. Miller et al.²⁴ suggested that the poor acceptability of IVR by older adults can be improved by designing IVR algorithms to better detect difficulties and adapt its function to fit individual needs during the interview process. To address older adults' hearing, speech, sight, cognition, and mobility issues, incorporating universal design in IVR is also necessary.²⁶ Based on our findings, we also suggest that giving consumers some practice of using IVR or using a live administered approach prior to IVR may help the consumers get familiar with the system. Furthermore, a more humanized and personal design with more lively interaction may enhance older adults' acceptance toward IVR and decrease operational challenges. There is also the potential to use newer technologies and approaches that can allow automated monitoring of function to take place but provide an experience that is more human-like and easier for older people to use. Virtual coaches, for example, are human-like animated computer characters with synthesized voices to engage in "conversations" with patients that can include counselling or assessment of health behaviors and outcomes.^{27, 28} Virtual coaches used on home computers or on mobile could be adapted to provide intensive and/or long-term functional monitoring.²⁹

This study has several limitations. First, the small sample size enlarges the range of confidence intervals for estimated reliability and reduces the estimation accuracy. Also, even though the recruited subjects were comparable with the racial and gender distribution of the population of the region, the sample was a convenience sample drawn from one urban area and may result in sample selection bias. In the future, the IVR-administered LLFDI-CAT can be tested in a more demographically diverse sample, in which the subgroup comparisons can also be made. Last, we used a brief telephone screening to enroll participants and made sure to exclude individuals with cognitive impairment, significant hearing loss or poor language comprehension that could limit their ability to accurately respond to questions, but this screening procedure was not conducted using standardized instruments. Despite these limitations, we believe this study provides valuable evidence support acceptable reliability of IVR administration of LLFDI-CAT.

CONCLUSION

LLFDI enables researchers and service providers to understand the function and disability of older adults. The combination of CAT and IVR technology further advanced the efficiency and effectiveness of administration. This study successfully applied IVR to administer the LLFDI-CAT and provides a preliminary test result in community dwelling older adults and demonstrates comparable reliability with the TIF version. Nevertheless, participants showed lower acceptability and familiarity with IVR comparing to TIF and administering the IVR interview first did lower the test retest reliability of the responses. To increase the usability of automated systems to monitor function, IVR, further investigation needs to address consumers' needs by improving the design of the administration system. This could be done by improving the design and protocols of IVR systems or adapting innovative technologies like Virtual Coaches to include functional monitoring components.

References

1. Jette AM, Haley SM, Coster WJ, et al. Late life function and disability instrument: I. development and evaluation of the disability component. *J Gerontol A Biol Sci Med Sci*. 2002; 57:M209–M216. [PubMed: 11909885]
2. Nagi, S. Disability concepts revisited: Implications for prevention. In: Pope, A.; Tarolov, A., editors. *Disability in America: Toward a National Agenda for Prevention*. Washington, DC: National Academy Press; 1991. p. 309-327.
3. World Health Organization. *International classification of functioning, disability and health (ICF)*. Geneva: World Health Organization; 2001.
4. Beauchamp M, Schmidt C, Pedersen M, et al. Psychometric properties of the late-life function and disability instrument: A systematic review. *BMC Geriatr*. 2014; 14:12. [PubMed: 24476510]
5. Jette AM, Haley SM, Ni P, et al. Creating a computer adaptive test version of the late-life function and disability instrument. *J Gerontol A Biol Sci Med Sci*. 2008; 63:1246–1256. [PubMed: 19038841]
6. Lee H, Friedman ME, Cukor P, et al. Interactive voice response system (IVRS) in health care services. *Nurs Outlook*. 2003; 51:277–283. [PubMed: 14688763]
7. McDaniel AM. Interactive voice response technology for outcomes monitoring. *Clin Nurse Spe*. 2004; 18:7–8.
8. Baer L, Jacobs DG, Cukor P, et al. Automated telephone screening survey for depression. *JAMA*. 1995; 273:1943–1944. [PubMed: 7783305]
9. Piette JD. Interactive voice response systems in the diagnosis and management of chronic disease. *Am J Manag Care*. 2000; 6:817–827. [PubMed: 11067378]
10. Kraft MR, Androwich I. Interactive voice response technology: A tool for improving healthcare. *Nurs Inform*. 2012:224.
11. Aiemagno SA, Cochran D, Feucht TE, et al. Assessing substance abuse treatment needs among the homeless: A telephone-based interactive voice response system. *Am J Public Health*. 1996; 86:1626–1628. [PubMed: 8916533]
12. Corkrey R, Parkinson L. Interactive voice response: Review of studies 1989–2000. *Behav Res Methods Instrum Comput*. 2002; 34:342–353. [PubMed: 12395550]
13. Greist JH, Mundt JC, Kobak K. Factors contributing to failed trials of new agents: Can technology prevent some problems? *J Clin Psychiatry*. 2002; 63 (Suppl 2):8–13. [PubMed: 15453008]
14. Brodey BB, Rosen CS, Brodey IS, et al. Reliability and acceptability of automated telephone surveys among spanish- and english-speaking mental health services recipients. *Ment Health Serv Res*. 2005; 7:181–184. [PubMed: 16194003]

15. Jette AM, Haley SM, Coster WJ, et al. Late life function and disability instrument: I. development and evaluation of the disability component. *J Gerontol A Biol Sci Med Sci*. 2002; 57:M209–M216. [PubMed: 11909885]
16. Haley SM, Jette AM, Coster WJ, et al. Late life function and disability instrument: II. development and evaluation of the function component. *J Gerontol A Biol Sci Med Sci*. 2002; 57:M217–M222. [PubMed: 11909886]
17. Friedman RH, Stollerman JE, Mahoney DM, et al. The virtual visit: Using telecommunications technology to take care of patients. *J Am Med Inform Assoc*. 1997; 4:413–425. [PubMed: 9391929]
18. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36): I conceptual framework and item selection. *Med Care*. 1992; 30:473–483. [PubMed: 1593914]
19. Schuck P. Assessing reproducibility for interval data in health-related quality of life questionnaires: Which coefficient should be used? *Qual Life Res*. 2004; 13:571–586. [PubMed: 15130022]
20. Portney, LG.; Watkins, MP. *Foundations of clinical research: Applications to practice*. 3. Mehalik, C., editor. New Jersey: Prentice Hall; 2008.
21. Chestnutt IG, Morgan MZ, Hoddell C, et al. A comparison of a computer-based questionnaire and personal interviews in determining oral health-related behaviours. *Community Dent Oral Epidemiol*. 2004; 32:410–417. [PubMed: 15541156]
22. Mundt JC, Bohn MJ, King M, et al. Automating standard alcohol use assessment instruments via interactive voice response technology. *Alcohol Clin Exp Res*. 2002; 26:207–211. [PubMed: 11964559]
23. Weiler K, Christ AM, Woodworth GG, et al. Quality of patient-reported outcome data captured using paper and interactive voice response diaries in an allergic rhinitis study: Is electronic data capture really better? *Ann Allergy Asthma Immunol*. 2004; 92:335–339. [PubMed: 15049397]
24. Miller DI, Bruce H, Gagnon M, et al. Improving older adults' experience with interactive voice response systems. *Telemed J E Health*. 2011; 17:452–425. [PubMed: 21631386]
25. Miller DI, Aube F, Talbot V, et al. Older people's attitudes toward interactive voice response systems. *Telemed J E Health*. 2014; 20:152–156. [PubMed: 24283251]
26. Dulude L. Automated telephone answering systems and aging. *Behav Inf Technol*. 2002; 21:171–184.
27. Bickmore TW, Caruso L, Clough-Gorr K, et al. It's just like you talk to a friend' relational agents for older adults. *Interact Comput*. 2005; 17:711–735.
28. Bickmore TW, Silliman RA, Nelson K, et al. A randomized controlled trial of an automated exercise coach for older adults. *J Am Geriatr Soc*. 2013; 61:1676–1683. [PubMed: 24001030]
29. Ellis T, Latham NK, DeAngelis TR, et al. Feasibility of a virtual exercise coach to promote walking in community-dwelling persons with parkinson disease. *Am J Phys Med Rehabil*. 2013; 92:472–481. [PubMed: 23552335]

Figure 1.
Likert rating of the participants' preference and difficulty of using the IVR and live-administered LLFDI-CAT

Table 1

Demographic characteristics of the participants at baseline (N=50)

Variable	IVR First Group (N=25)	TIF First Group(N=25)
Age (years), mean (SD)	78.5 (7.6)	76.4 (8.1)
Female, number (%)	19 (76)	19 (76)
PF-10, mean (SD)	57.7 (24.5)	54.6 (24.9)
Years since the most recent hip fracture, mean (SD)	2.8 (1.4)	4.2 (4.6)

Table 2
Test-retest reliability and concordance (baseline to 1-week comparison) of IVR and TIF version of LLFDI-CAT

	All cases ICCs (95% CI)		IVR First Group ICCs (95% CI)		TIF First Group ICCs (95% CI)	
	Disability	Function	Disability	Function	Disability	Function
	Test-retest reliability					
IVR	0.80 (0.68, 0.88)	0.79 (0.65, 0.87)	0.65 (0.33, 0.83)	0.54 (0.18, 0.77)	0.94 (0.86, 0.97)	0.97 (0.93, 0.99)
TIF	0.74 (0.58, 0.84)	0.85 (0.75, 0.91)	0.88 (0.74, 0.95)	0.70 (0.43, 0.86)	0.66 (0.38, 0.83)	0.95 (0.9, 0.98)
	Concordance					
Baseline	0.74 (0.58, 0.84)	0.83 (0.72, 0.9)	0.84 (0.67, 0.93)	0.67 (0.37, 0.84)	0.69 (0.43, 0.85)	0.97 (0.92, 0.98)
1-Week Follow-up	0.96 (0.93, 0.98)	0.97 (0.95, 0.98)	0.95 (0.88, 0.98)	0.98 (0.95, 0.99)	0.96 (0.91, 0.98)	0.97 (0.93, 0.99)