

Differential Diagnoses That Final-Year Medical Students Need to Consider: A Modified Delphi Study

Yuka Urushibara-Miyachi (✉ yuka.urushibara@gmail.com)

Kyoto University

Makoto Kikukawa

Kyushu University

Masatomi Ikusaka

Chiba University Hospital

Junji Otaki

Tokyo Medical University

Hiroshi Nishigori

Nagoya University

Research Article

Keywords: Clinical reasoning, Contrastive learning, Undergraduate medical education, Core curriculum, Expert consensus, Modified Delphi approach

Posted Date: December 21st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-127514/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at BMC Medical Education on April 23rd, 2021. See the published version at <https://doi.org/10.1186/s12909-021-02652-5>.

Abstract

Background: Contrastive learning is known to be effective in teaching medical students how to generate diagnostic hypotheses in clinical reasoning. However, there is no consensus on comprehensive lists of differential diagnoses across different medical disciplines regarding the common symptoms that should be learned as part of the undergraduate medical curriculum. In Japan, the national model core curriculum for undergraduate medical education was revised in 2016, and lists of differential diagnoses for 37 common symptoms were introduced into the curriculum. This study aimed to validate the list of items based on expert consensus for use as a reference worldwide.

Methods: The authors used a modified Delphi method to develop consensus among a panel of 23 expert physician-teachers in clinical reasoning from across Japan. The panel evaluated the items on a 5-point Likert scale, based on whether a disease should be hypothesized by final-year medical students considering a given symptom. They also added other diseases that should be hypothesized. A positive consensus was defined as both a 75% rate of panel agreement and a mean of 4 or higher with a standard deviation of less than 1 on the 5-point scale. The study was conducted between September 2017 and March 2018.

Results: This modified Delphi study identified 275 essential and 67 supplemental items corresponding to the differential diagnoses for 37 common symptoms that Japanese medical students should master before graduation.

Conclusions: The lists developed in the study can be useful for teaching and learning how to generate initial hypotheses by encouraging students' contrastive learning. Although the lists may be specific to the Japanese context, the lists and process of validation are generalizable to other countries for building national consensus on the content of medical education curricula.

Background

Amid the growing complexity of healthcare settings and considering the social accountability of physicians in ensuring patient safety, medical schools are required to educate and evaluate medical students with sufficiently high standards, so that they become physicians who make fewest possible diagnostic errors. This is particularly evident in the worldwide trend of incorporating the assessment of clinical reasoning skills in high-stakes examinations such as national medical licensing exams, as represented by Step 2, "Clinical Skills," of the United States Medical Licensing Examination (USMLE) [1]. Thus, greater educational support is required for students to acquire the competence to anticipate a set of differential diagnoses from the earliest phase of the diagnostic process, gather confirming and refuting information according to an initial hypothesis, select and perform the relevant history taking and physical examination, and interpret the findings to confirm or deny the initial hypothesis.

In this context, the lack of development of diagnostic hypotheses remains an issue in the teaching of clinical reasoning to medical students. For example, medical students learn how to take a patient's

history without anticipating differential diagnoses, even though the literature suggests that diagnostic errors can be reduced by querying an initial hypothesis [2]. Furthermore, attempting to diagnose without generating a hypothesis may reduce students' reasoning performance [3]. Nevertheless, many medical schools have taught physical examination maneuvers in isolation, usually following a systematic "head-to-toe" approach [4]. Recent studies in cognitive load theory have described that such fragmented reasoning may lead to diagnostic errors [5].

How can teachers effectively instruct medical students in hypothesis generation? Schmidt and Mamede [6] found that there is no evidence elucidating the most effective method for teaching clinical reasoning to medical students. In the highly specialized wards of teaching hospitals, which feature increasingly shorter patient stays, patient interaction with and exposure to role model doctors in clinical clerkships may be insufficient for students to gain the competence required in clinical reasoning. This is particularly the case regarding the need to consider the full range of differential diagnoses across all medical disciplines and specialties [7]. Students usually learn specialty-specific or disease-oriented reasoning skills from specialists, who tend to generate diagnostic hypotheses focused on a particular organ system [8]. Variation among the patient cases that students experience tends to be limited, and feedback from attending doctors is opportunistic [6]. Instead, various pre-clinical curricula have been introduced to teach students clinical reasoning, in which cases are usually also limited [6]. In this context, previous research has suggested that a "comparing and contrasting" approach is effective for medical students to foster illness scripts in their minds [6, 9]. Medical students without sufficient clinical experience can effectively formulate an illness script of disease by comparing and contrasting the discriminating clinical features of other competing diagnoses in terms of a particular symptom.

Nevertheless, in the context of undergraduate medical education, there was no learning resource supporting medical students' contrastive learning by covering a wider variety of signs and symptoms. As of 2017, when this study was conducted, no previous study had developed a consensus on comprehensive lists of differential diagnoses across different medical disciplines regarding the common symptoms to be learned during the undergraduate medical curriculum. Therefore, this research aimed to develop lists of a limited number of diagnostic considerations regarding all the signs and symptoms that medical students should learn before graduation. The emerging lists will be beneficial for developing both case scenarios that consider a plausible set of differential diagnoses suitable for medical students' reasoning competence and assessment tools for checking differentiating information in the process of history taking and physical examination.

Although these lists can be universally applicable in the context of undergraduate medical education, they may also be specific to the local social context in which they are developed and introduced. The reason for this specificity is that various epidemiological factors and societal needs can influence what diseases medical students learn to diagnose within their countries. In Japan, the fourth version of the national core curriculum for undergraduate medical education in 2016 newly introduced lists of differential diagnoses regarding 37 common symptoms (Table 1) that ought to be learned as part of the six-year undergraduate curriculum [10]. An original set of lists was developed through a review of the

previous literature on clinical reasoning by committee members consisting of general internists specializing in teaching clinical reasoning, followed by a revision based on public feedback. As this process possibly reflected the personal perspectives of a limited number and range of specialists with authority, we attempted to validate the lists by building a consensus among experts in clinical reasoning education through a systematic, iterative process. Thus, the research question for this study asked: what are the differential diagnoses that final-year medical students need to consider for the symptoms listed in the national guideline for undergraduate medical education?

Table 1

The 37 common symptoms in the national model core curriculum for undergraduate medical education (revised in 2016)

1. Fever	14. Hemosputum/hemoptysis	27. Lymphadenopathy
2. General fatigue	15. Dyspnea	28. Abnormality of urine and urination
3. Appetite loss	16. Chest pain	29. Hematuria/proteinuria
4. Weight gain/loss	17. Palpitation	30. Menstrual disorders
5. Shock	18. Pleural effusion	31. Anxiety/depression
6. Heart arrest	19. Dysphagia	32. Memory loss
7. Disturbance of consciousness/syncope	20. Abdominal pain	33. Headache
8. Seizure	21. Nausea/vomit	34. Motor paralysis
9. Dizziness	22. Hematemesis/melena	35. Back pain
10. Dehydration	23. Constipation/diarrhea	36. Arthralgia/swollen joint
11. Edema	24. Jaundice	37. Trauma/burn
12. Rash	25. Abdominal distension/mass	
13. Cough/sputum	26. Anemia	

Methods

In this study, we utilized a modified Delphi method [11, 12]. In the original Delphi method [13], an initial list to be examined is established based on feedback from experts in the first round. However, in the modified Delphi method, the initial list is mostly produced by the researchers based on a literature review, interviews with relevant professionals, or other academic methods [14]. The Delphi and modified Delphi methods allow researchers to gather and achieve consensus on the opinions of experts through a series of structured questionnaires, which are conducted anonymously to avoid the influence of authority among the experts [13]. This is considered one of the strengths of these methods, particularly in East-

Asian countries such as Japan, where hierarchical social relationships tend to have a strong influence on the stakeholders of decision-making processes [15].

Both methods have been used in a variety of studies to establish a consensus on core competencies or curricula regarding a specific topic or domain among medical specialties. For instance, Alahlafi and Burge [16] conducted a Delphi study to build a consensus on what medical students need to learn about psoriasis. Battistone, Barker, Beck, Tashjian, and Cannon [17] defined the core skills of shoulder and knee assessment using the Delphi method. Moore and Chalk [18] used the Delphi method to produce a consensus on neurological physical examination skills that should be learned by medical students. Finally, Moercke and Eika [19] used the Delphi method to identify the required clinical skills and minimum competency levels therein during undergraduate training.

The expert panel

Although there is no consensus on the most appropriate number of experts to include in a panel for Delphi studies, previous studies have generally required at least 20 expert participants for sufficient reliability [20]. Considering the average response rate in past Delphi studies of approximately 80% [21], we aimed to recruit 25 participants. The previous literature also suggested that recruiting a variety of participants may produce higher quality results that are more generally acceptable [15]. Thus, we used purposeful sampling to recruit participants from different areas across Japan and different types of institutions ranging from community hospitals to university hospitals. The four authors (MI, JO, MK, HN), who were general internists specializing in teaching clinical reasoning, produced a list of 23 candidates for discussion. YM, who was a physician-researcher on clinical reasoning, contacted all the potential participants via email to ensure their interest and obtain their agreement to participate; all the candidates agreed to join our research. Informed consent was obtained from all participants. Thus, we assembled a panel of 23 physicians (clinical faculty members of medical schools or clinical teachers in teaching hospitals) with more than ten years of experience in the practice and teaching of clinical reasoning to medical students and with an understanding of the national model core curriculum.

The initial lists

The initial lists regarding the 37 common symptoms consisted of 277 items. To make each list sufficient but as short as possible for the study, the four authors decided to incorporate only the 170 diseases designated as the minimum requirement in the guidelines for the Japanese national licensing examination [22] into the lists. The questionnaire, edited by YM, was piloted by two members of the research team (HN and MK). As previous studies, including Hasson et al. [13], suggested two or three rounds for Delphi studies, we designed our study using a two-round modified Delphi method, with an optional additional third round for re-evaluating certain items, an option that we ended up using. In all rounds, questionnaires with the lists to be evaluated were provided to the participants via a web-based survey system (Google Forms: available at <https://www.google.com/intl/ja/forms/about/>).

In Round 1, for each list of differential diagnoses for the 37 symptoms, the participants were asked to evaluate on a 5-point Likert scale (1 = absolutely needs to be excluded, 5 = absolutely needs to be included) whether each item (i.e., disease) should be hypothesized by final-year medical students when diagnosing a patient with the given symptom. The participants were also asked to add any other diseases they considered relevant to include in the list. Their responses were anonymously collected and analyzed with a predefined standard for positive consensus as follows. Based on our literature review, which included research from Dielissen et al. [21] and Heiko [23], our standard was 1) a mean score of 4 and higher with 2) a standard deviation of less than 1 and 3) 75% or more of the experts scoring 4 or 5 (i.e., 75% agreement). The participants were given two weeks to evaluate the lists in each round. To keep the response rates sufficiently high, the participants received reminder e-mails a few days before each deadline. Two researchers (YM and HN) analyzed the primary data and discussed the results with the other researchers. Among all the additional items suggested by the panel, only the diseases listed as minimum requirements for the national licensing exam were incorporated into the revised list for Round 2.

The second version of the list, to be evaluated in Round 2, consisted of 1) diseases from the original lists that met our positive consensus standard and 2) diseases added by the panelists that were part of the minimum requirements for the national licensing exam. Additionally, the panel members re-evaluated diseases from the initial lists that did not meet the standard to ensure the appropriateness of their exclusion in comparison with the suggested and newly added diseases. The panel provided free-form comments on the questionnaire in every round. The study took place between September 2017 and March 2018 and was approved by the institutional research board of the Kyoto University Graduate School of Medicine (R0481). This study complied with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidelines [24].

Results

The participating experts, who had 11 years or more (average 21 years) of clinical expertise and more than 5 years (average 13 years) of teaching expertise, were all male and had a variety of backgrounds, including general medicine/general internal medicine (n = 20), family medicine (n = 3), and internal medicine (n = 2); eight of them had more than two specialties (Table 2).

Table 2
Demographics of the expert panel (n = 23)

Characteristic	No. (%)
Years in practice	14 (61)
10–19 years	4 (17)
20–29 years	5 (22)
More than 30 years	
Years of involvement in clinical teaching	6 (26)
Less than 10 years	13 (57)
10–19 years	4 (17)
More than 20 years	
Certifications/specialties (Multiple answers allowed)	20 (88)
General Medicine/General Internal Medicine	3 (13)
Family Medicine	2 (9)
Internal Medicine	1 (4)
Geriatrics	1 (4)
Emergency Medicine	1 (4)
Gastroenterology	1 (4)
Neurology	1 (4)
Cardiology	2 (9)
Medical Education	8 (35)
More than two specialties	
Practice/educational settings	19 (83)
University/University hospital	4 (17)
Community hospital	

Characteristic	No. (%)
Regions	9 (39)
Kanto	5 (22)
Chubu	2 (9)
Kansai	1 (4)
Chugoku	6 (26)
Kyushu	

In the first round, 47 items were eliminated from the initial lists (of 277 items) according to the pre-set standard. Among the 428 items that the study participants additionally suggested, 185 items that were also part of the minimum requirements for the national licensing examination were included in the revised list for Round 2 (Fig. 1).

After analyzing the feedback our panel members gave in Round 2, we examined the face validity of all the lists on which consensus had been reached. This third version consisted of 13 items on average per symptom (and a median of 7 items) as essential differential diagnoses for final-year medical students. In total, the final lists comprised 275 items; 187 items, including 79 items from the initial lists, were eliminated in the two rounds of the study.

In their free comments, some of the experts questioned the validity of the minimum requirement for the national licensing exam being used as part of the standard. We decided to conduct one additional round for our panelists to evaluate the diseases they suggested in Round 1 that were not part of the minimum requirements for the national licensing exam but were included elsewhere in the guideline. Among the 257 items that were added by the experts in Round 1, 89 were eliminated because they were not identified in the guidelines. Among 167 items included in the guideline, 67 items that met the consensus standard were incorporated in supplemental lists.

The final lists consisted of essential differential diagnoses, and the supplemental lists are available as Additional file 1. As an example, the list of essential differential diagnoses for the symptom of chest pain as well as the supplemental list are illustrated in Table 3. Of the 23 participants who took part in the study, 22 completed the first and second rounds (96%), and 20 completed the additional round (87%).

Table 3
The example of “Chest pain”

Essential differential diagnoses	Mean score (SD)	Number of experts who chose “must include” (%)
Respiratory	4.6 (0.57)	21 (95)
Pulmonary embolism	4.7 (0.47)	22 (100)
Pneumothorax	4.9 (0.29)	22 (100)
Cardiovascular	4.9 (0.29)	22 (100)
Acute coronary syndrome	4.8 (0.39)	22 (100)
Acute aortic dissection	4.3 (0.75)	20 (91)
Rupture of aortic aneurysm		
Psychogenic		
Panic disorder		
Supplemental diagnoses	Mean score (SD)	Number of experts who chose “must include” (%)
Acute pericarditis	4.1 (0.87)	15 (75)
Pleurisy	4.4 (0.57)	19 (95)
Herpes Zoster	4.0 (0.95)	15 (75)
Abbreviations: SD, Standard Deviation		

Discussion

This modified Delphi study identified 275 essential and 67 supplemental items as the differential diagnoses for 37 common symptoms that Japanese medical students should master before graduation. The lists developed in the study can be useful for teaching and learning how to generate initial hypotheses because, as Bowen and ten Cate [9] suggest, a preselection of differential diagnoses to consider is crucial for novice learners acquiring this competence. This method could also be transferable to other countries with a similar growing emphasis on patient safety and de-emphasis on hospital stay [25]. Moreover, both the lists and method can be useful in other countries when developing a national consensus on lists of symptom-specific differential diagnoses across all medical disciplines.

The knowledge and skills regarding essential diseases for final-year medical undergraduates should be equivalent to those of new graduates. In the move toward competency-based medical education, Japan has been reforming competency-based postgraduate programs per the competencies developed during undergraduate medical education [26]. Thus, this research is significant because clinical faculty members and clinical teachers with experience in teaching both medical students and residents participated in the

study. The results of our study can contribute to further revision of the national guideline for undergraduate medical education as well as stronger continuity between undergraduate and postgraduate education.

One of the study's strengths is the high response rate across all three rounds of the Delphi process. According to Cantrill, Sibbald, and Buetow [27], response rates can influence the validity of the derived consensus. The high response rate may also indicate the perceived importance of this study to our panelists. Considering the limited reduction in respondent numbers after every round, we would argue that the results are valid. The validity of our study was also enhanced by ensuring a representative panel of experts recruited from a variety of institutions from community hospitals to university hospitals across Japan.

This study also had some limitations. First, we limited the differential diagnoses to those listed in the minimum requirements of the guideline for the national licensing examination and adopted a categorical representation of diseases according to the guideline. For example, when the panelists added "renal failure," this was converted into "acute kidney disease" and "chronic kidney failure" as indicated in the guideline, and both were evaluated independently in the following rounds. Although the authors were aware of the possibility that this might eliminate some diseases that should be actively considered by final-year medical students, we opted to make the lists sufficiently compact with a low variance in terminology to avoid "curriculum hypertrophy" [28]. To reduce the possible influence of this limitation, we designed an additional round to evaluate items outside of the minimum essentials. While this study attempted to produce a consensus on the minimum requirements for final-year medical students, it does not imply that the generated lists are thorough or complete.

Second, some of the symptoms and their classifications might have confused our panelists. For instance, among the 37 common symptoms considered in the model core curriculum, two different clinical conditions were combined into one condition several times, such as "Hematuria" and "Proteinuria," "Anxiety" and "Depression," and "Disturbance of consciousness" and "Syncope." Some of the experts claimed that these were clinically distinct due to different but partly overlapping working diagnoses. Moreover, the panelists pointed out that the relationship between some independent symptoms among the 37 was ambiguous, such as "Vertigo" and "Syncope." They also reported a few irregular cases in which the concept of "clinical reasoning" did not seem to apply, such as cases of "Trauma" and "Burns." These notes are crucial for future improvement of both the core medical curriculum and the guidelines for the national licensing examination.

Finally, initial differential diagnoses for a symptom can be dependent on a patient's sex and age, neither of which were considered in the lists generated in this study. We would advise future users of the lists to consider this background information.

Conclusions

In conclusion, our study can contribute to promoting the integration of fragmented clinical reasoning teaching processes by encouraging students' contrastive learning as well as ensuring greater continuity between undergraduate and postgraduate medical education in the Japanese setting. This is also important for other countries where the ever-expanding competencies required of medical students exacerbate the fragmentation of medical education [5]. Furthermore, as the use of artificial intelligence (AI) in the teaching of clinical reasoning has been increasing [29], the results of this study may be useful as a guideline regarding which diagnoses should be considered in machine learning. Similarly, the findings can also be beneficial in the development of clinical decision support systems. The method and its potential impact on future AI-assisted medical education could apply to other countries. For our next step, we aim to develop a list of which medical history and physical examination elements final-year medical students are required to gather and master when considering the differential diagnoses on the lists resulting from this study.

Declarations

Ethics approval and consent to participate

The study protocol was performed in accordance with the Declaration of Helsinki, reviewed and approved by the institutional research board of the Kyoto University Graduate School of Medicine (R0481, 6/9/2017).

Consent for publication

Not applicable

Availability of data and materials

The datasets used and analyzed during the current study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study is supported by the Pfizer Health Research Foundation. The sponsor was not involved in the study design; the collection, analysis, or interpretation of the data; the preparation of the manuscript; or the decision to submit it for publication.

Authors' contributions

The four authors (MI, JO, MK, HN) produced a list of 23 candidates and discussed it with each other. YM contacted all the potential participants via email to ensure their interest and obtain their agreement to

participate. The questionnaire, edited by YM, was piloted by HN and MK. YM and HN analyzed the primary data and discussed the results with the other researchers. All authors read and approved the final manuscript.

Acknowledgements

Not applicable

Authors' information

Yuka Urushibara-Miyachi, MD, MHPE, MSc, is a family physician and part-time lecturer with the Faculty of Medicine, Kyoto University, Kyoto, Japan.

Makoto Kikukawa, MD, MMed, PhD, is a lecturer with the Department of Medical Education, Faculty of Medical Sciences at Kyushu University, Fukuoka, Japan.

Masatomi Ikusaka, MD, PhD, is a professor in the Department of General Medicine at Chiba University Hospital, Chiba, Japan.

Junji Otaki, MD, DMedSc, is a professor at the Center for Medical Education, Faculty of Medicine, Tokyo Medical University, Tokyo, Japan.

Hiroshi Nishigori, MD, MMed, PhD, is a professor at the Center for Medical Education, Graduate School of Medicine, Nagoya University, Nagoya, and a Visiting Project Leader Professor at the Medical Education Center, Graduate School of Medicine, Kyoto University, Kyoto, Japan.

References

1. Swanson DB, Roberts TE. Trends in national licensing examinations in medicine. *Med Edu.* 2016;50(1):101–14.
2. Coderre S, Wright B, McLaughlin K. To think is good: querying an initial hypothesis reduces diagnostic error in medical students. *Acad Med.* 2010;85(7):1125–9.
3. Norman GR, Brooks LR, Colle CL, Hatala RM. The benefit of diagnostic hypotheses in clinical reasoning: experimental study of an instructional intervention for forward and backward reasoning. *Cogn Instr.* 1999;17(4):433–48.
4. Yudkowsky R, Otaki J, Lowenstein T, Riddle J, Nishigori H, Bordage G. A hypothesis-driven physical examination learning and assessment procedure for medical students: initial validity evidence. *Med Educ.* 2009;43(8):729–40.
5. Young JQ, Van Merriënboer J, Durning S, Ten Cate O. Cognitive load theory: implications for medical education: AMEE Guide No. 86. *Med Teach.* 2014;36(5):371–84.
6. Schmidt HG, Mamede S. How to improve the teaching of clinical reasoning: a narrative review and a proposal. *Med Edu.* 2015;49(10):961–73.

7. Rencic J, Trowbridge RL, Fagan M, Szauter K, Durning S. Clinical reasoning education at US medical schools: results from a national survey of internal medicine clerkship directors. *J Gen Intern Med.* 2017;32(11):1242–6.
8. Hashem A, Chi MT, Friedman CP. Medical errors as a result of specialization. *J Biomed Inform.* 2003;36(1–2):61–9.
9. Bowen JL, ten Cate O. Prerequisites for learning clinical reasoning. In: ten Cate O, Custers E, Durning S, editors. *Principles and practice of case-based clinical reasoning education. Innovation and change in professional education*, vol. 15. Springer; 2018.
10. Ministry of Education, Culture, Sports, Science and Technology. *Igaku Kyoiku Moderu Koa Karikyuramu (Heisei 28 nendo kaiteiban), Shigaku Kyoiku Moderu Koa Karikyuramu (Heisei 28 nendo kaiteiban) no kohyo ni tsuite [Regarding the model core curriculum for medical education (2016 version) and the model core curriculum for dental education (2016 version).]* https://www.mext.go.jp/b_menu/shingi/chousa/koutou/033-2/toushin/1383962.htm. Published March 31, 2017. Accessed November 15, 2020. [in Japanese]
11. McKenna HP. The Delphi technique: a worthwhile research approach for nursing? *J Adv Nurs.* 1994;19(6):1221–5.
12. Newman LR, Lown BA, Jones RN, Johansson A, Schwartzstein RM. 2009. Developing a peer assessment of lecturing instrument: lessons learned. *Acad Med.* 84(8):1104–10.
13. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. *J Adv Nurs.* 2000;32(4):1008–15.
14. Custer RL, Scarcella JA, Stewart BR. The modified Delphi technique - A rotational modification. *CTE J.* 1999;15(2):50–8.
15. Kikukawa M, Stalmeijer RE, Emura S, Roff S, Scherpbier AJ. An instrument for evaluating clinical teaching in Japan: content validity and cultural sensitivity. *BMC Med Educ.* 2014;14(1).
16. Alahlafi A, Burge S. What should undergraduate medical students know about psoriasis? Involving patients in curriculum development: modified Delphi technique. *BMJ.* 2005;330(7492):633–6.
17. Battistone MJ, Barker AM, Beck JP, Tashjian RZ, Cannon GW. Validity evidence for two objective structured clinical examination stations to evaluate core skills of the shoulder and knee assessment. *BMC Med Educ.* 2017;17(1):13.
18. Moore FG, Chalk C. The essential neurologic examination: what should medical students be taught? *Neurol.* 2009;72(23):2020–23.
19. Moercke AM, Eika B. What are the clinical skills levels of newly graduated physicians? Self-assessment study of an intended curriculum identified by a Delphi process. *Med Educ.* 2002;36(5):472–8.
20. Dunn WR, Hamilton DD, Harden RM. Techniques of identifying competencies needed of doctors. *Med Teach.* 1985;7(1):15–25.
21. Dielissen P, Verdonk P, Bottema B, Kramer A, Lagro-Janssen T. Expert consensus on gender criteria for assessment in medical communication education. *Patient Educ Couns.* 2012;88(2):189–95.

22. Ministry of Health, Labour and Welfare. Health Policy Bureau. Medical Professions Division. Heisei 30 Nendo Ishi Kokkashiken Shutsudai Kijun ni tsuite [Regarding the guideline for national licensure examination 2018 version.]. Mhlw.go.jp. <https://www.mhlw.go.jp/stf/shingi2/0000128981.html>. Published June 30, 2016. Accessed November 15, 2020. [in Japanese]
23. Heiko A. Consensus measurement in Delphi studies: review and implications for future quality assurance. *Technol Forecast Soc Change*. 2012;79(8):1525–36.
24. Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007;147(8):573–7.
25. Peters M, ten Cate O. Bedside teaching in medical education: a literature review. *Perspect Med Educ*. 2014;3(2):76–88.
26. Konishi et al. Japan Society for Medical Education, Postgraduate Medical Education Committee. Present undergraduate medical education with connection to Postgraduate education. *Med Educ*. 2017;48(6):387–94. [in Japanese]
27. Cantrill JA, Sibbald B, Buetow S. The Delphi and nominal group techniques in health services research. *Int J Pharm Pract*. 1996;4(2):67–74.
28. Abrahamson S. Diseases of the curriculum. *J Med Educ*. 1978;53(12):951–7.
29. Chan KS, Zary N. Applications and challenges of implementing artificial intelligence in medical education: integrative review. *JMIR Med Educ*. 2019;5(1):e13930.

Figures

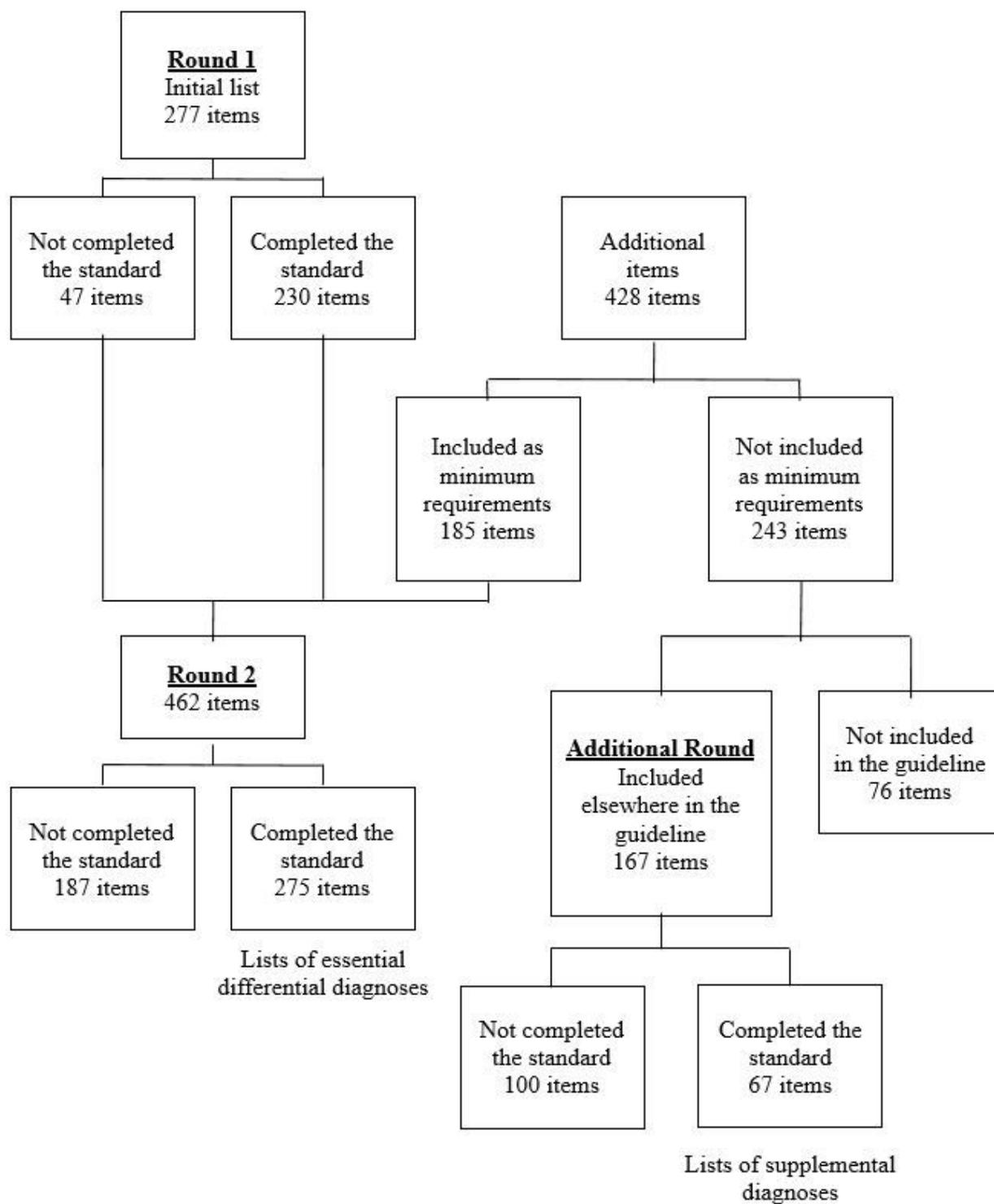


Figure 1

Numbers of items in the three rounds

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.docx](#)