



**HAL**  
open science

# Évaluation de la fragilité en oncologie gériatrique : développement et validation d'une nouvelle échelle de dépistage.

Claudia Tapia Martinez

## ► To cite this version:

Claudia Tapia Martinez. Évaluation de la fragilité en oncologie gériatrique : développement et validation d'une nouvelle échelle de dépistage.. Santé publique et épidémiologie. Université Paris-Saclay, 2017. Français. NNT : 2017SACLS368 . tel-01921658

**HAL Id: tel-01921658**

**<https://theses.hal.science/tel-01921658>**

Submitted on 14 Nov 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Assessment of frailty in geriatric oncology: development and validation of a new screening tool

## *Evaluation de la fragilité en oncologie gériatrique : développement et validation d'une nouvelle échelle de dépistage.*

THESE DE DOCTORAT DE L'UNIVERSITE PARIS-SACLAY  
PREPAREE A L'UNIVERSITE PARIS SUD

ÉCOLE DOCTORALE N°570 : Santé publique (EDSP)  
Spécialité de doctorat: sante publique - épidémiologie

Sous la direction du Pr. Sylvie BASTUJI-GARIN et du Pr. Etienne AUDUREAU

Thèse présentée et soutenue à Créteil, le 13 novembre 2017, par

**Claudia Tapia**

### Composition du Jury :

Dr. Carine BELLERA Biostatisticienne, Université de Bordeaux	Rapporteur
Pr. Olivier GUERIN PU-PH, Université de Nice Sophia-Antipolis	Rapporteur
Pr. Jean Luc NOVELLA PU-PH, Université de Reims Champagne-Ardenne	Président
Pr. Lydia GUITTET PU-PH, Université de Caen Basse Normandie	Examineur
Dr. Marie HERR MCU-PH, Université Paris Saclay	Examineur
Pr. Sylvie BASTUJI-GARIN PU-PH, Université Paris-Est Créteil	Directeur de thèse
Pr. Etienne AUDUREAU PU-PH, Université Paris-Est Créteil	Co-Directeur de thèse



**Titre : EVALUATION DE LA FRAGILITE EN ONCOLOGIE GERIATRIQUE : DEVELOPPEMENT ET VALIDATION D'UNE NOUVELLE ECHELLE DE DEPISTAGE.**

**Mots clés :** fragilité, sujet âgé, cancer, dépistage, développement, validation, analyse de survie.

**Résumé :**

**Contexte :** Une évaluation gériatrique (EG) est recommandée pour tous les patients âgés cancéreux afin d'identifier d'éventuels problèmes de santé et ainsi optimiser la stratégie thérapeutique. Néanmoins, cette évaluation est longue et requiert une expertise spécifique. Plusieurs échelles simples ont été proposées comme outils de dépistage mais 1) aucune ne dispose des propriétés diagnostiques adéquates pour détecter la fragilité, 2) ces échelles ont le plus souvent été développées sur la base exclusive d'opinions d'experts sans développement statistique psychométrique spécifique, et enfin 3) la définition du concept de « fragilité » varie selon les échelles et peu de données permettent d'apprécier leur robustesse vis-à-vis de variations du gold standard. Le G8 est l'instrument de dépistage le plus récent mais sa spécificité reste perfectible. Par conséquent, l'objectif de cette thèse était de : 1) développer et valider un outil de dépistage performant de la fragilité, 2) évaluer sa robustesse vis-à-vis de changements du gold standard, et 3) évaluer sa validité pronostique vis-à-vis de la mortalité.

**Méthodes :** Les données étaient issues de deux cohortes prospectives de patients âgés atteints de cancer : ELCAPA (2007-2012: n=729 [développement] ; ONCODAGE (2008-2010: n=1304 [validation externe]). Une EG anormale était définie par  $\geq 1$  domaine altéré sur 7 tests validés (ADL, IADL, MMSE, Mini-GDS [ELCAPA] ou GDS-15 [ONCODAGE], MNA, CIRS-G et TUG). La procédure de développement incluait: analyse des correspondances multiples, régression logistique multivariée, validation interne (bootstrap). Les sensibilité, spécificité, aire sous la courbe ROC (AUROC) ont été calculées pour chaque cohorte et définition de la fragilité (gold standard) afin d'analyser la robustesse aux variations de gold standard ( $\geq 1$  ou  $\geq 2$  anomalies à l'EG;  $\geq 1$  intervention gériatrique; profil de fragilité par approche en classes latentes et deux classifications basées sur avis d'experts). Des modèles pronostiques de survie globale ont été construits à 1 et 3 ans après diagnostic de cancer : modèles de Cox à risques proportionnels univariés, puis ajustés sur l'âge, le sexe, le traitement anticancéreux et le statut métastatique, et analyses stratifiées sur le statut métastatique et le site tumoral.

**Résultats :** Une version modifiée du G8 comprenant 6 items a été développée : perte de poids, problèmes neuropsychologiques, statut fonctionnel, état de santé perçu, poly-prescription ( $\geq 6$  médicaments/jour) et existence d'une insuffisance cardiaque/coronaropathie. La sensibilité était de 89,2% (intervalle de confiance à 95% : 86,5-91,5), spécificité 79,0% (69,4-86,6) et AUROC 91,6% (89,3-93,9) avec de meilleures propriétés diagnostiques en validation interne (ELCAPA) et externe (ONCODAGE) par rapport au G8 original. Pour l'analyse de robustesse, quel que soit le gold standard, les AUROC étaient  $\geq 80\%$  pour les deux outils, mais significativement plus élevées en faveur du G8 modifié pour la prédiction de 4 des 6 gold standard testés. Les deux scores anormaux étaient indépendamment associés à la survie globale à 1 an : hazard ratio ajusté [HRa]=4,3[G8]/4,9[G8 modifié] et à 3 ans : HRa=2,9/2,6. Les associations persistaient après stratification sur le statut métastatique et dans la plupart des sites cancéreux. Pour les deux scores, la construction de classes de risque croissant montraient une relation graduée avec la mortalité.

**Conclusions :** Le développement d'une version modifiée de l'outil de dépistage G8 qui ne comporte que six items à recueillir a permis d'améliorer la performance diagnostique de l'instrument avec une meilleure homogénéité selon la localisation du cancer. Ces caractéristiques ont été confirmées dans une large cohorte prospective multicentrique de patients âgés atteints de cancer, démontrant l'intérêt de l'outil modifié pour faciliter la sélection des patients pouvant bénéficier d'une EG. Les deux outils ont démontré une robustesse aux changements du gold standard, avec une performance diagnostique globalement supérieure du G8 modifié pour détecter une large palette de profils de santé évocateurs de fragilité. Les deux scores ont enfin démontré une valeur pronostique forte pour la survie globale, quel que soit le statut métastatique ou le site de la tumeur. Nos résultats renforcent l'utilité clinique de ces instruments dans le cadre de l'oncologie gériatrique.

**Title : ASSESSMENT OF FRAILTY IN GERIATRIC ONCOLOGY: DEVELOPMENT AND VALIDATION OF A NEW SCREENING TOOL**

**Key words:** frailty, older person, cancer, development, validation, survival analysis.

**Abstract:**

**Background:** The comprehensive geriatric assessment (GA) is recommended in older cancer patients to inventory health problems and tailor treatment decisions accordingly, but is time- and resource-consuming. Several shorter scales have been proposed as screening tools but 1) their diagnostic performance is insufficient, 2) most instruments have been developed exclusively on the basis of experts' opinions without specific statistical/psychometric development, and 3) a wide variability of criteria have been used to define "frailty" as the gold standard, with no investigations of their influence on the diagnostic properties and/or concepts being actually measured. The G8 is the most recent screening tool but its specificity still leaves room for improvement. Therefore, the objective of this thesis was to 1) develop and validate a new screening tool for frailty that achieves high diagnostic accuracy, 2) evaluate its robustness to modifications on the reference definition, and 3) assess its prognostic value for survival.

**Methods:** We used the ELCAPA prospective cohort of cancer patients aged  $\geq 70$  years referred to geriatricians for GA (2007–2012:  $n=729$  [training set]) and the ONCODAGE multicenter prospective cohort (2008–2010:  $n=1304$  [external validation set]). An abnormal GA was defined as  $\geq 1$  impaired domain across 7 validated tests: ADL, IADL, MMSE, Mini-GDS (ELCAPA) or GDS-15 (ONCODAGE), MNA, CIRS-G, and TUG. Multiple correspondence analysis, multivariate logistic regression, and bootstrapped internal validation were performed sequentially. Sensitivity, specificity and the area under the ROC curve (AUROC) were calculated for each set and for each gold standard definition tested (detection of  $\geq 1$  or  $\geq 2$  impaired GA components;  $\geq 1$  geriatric intervention; a vulnerable profile as defined by a latent class approach or by expert-based classifications). Cox proportional-hazards models were built to assess the predictive value of abnormal G8 and modified G8 scores in the ELCAPA cohort (2007–2014:  $n=1333$ ). Endpoints were overall 1- and 3-years survival. Sensitivity analyses involved adjusting for age, gender, treatment, metastasis and tumor site and stratifying by tumor site and metastatic status.

**Results:** An optimized version of the G8 was developed including six items: weight loss, cognition/mood, performance status, self-rated health status, polypharmacy ( $\geq 6$  medications/day), and heart failure/coronary heart disease. For the original G8, sensitivity was 87.2% (95% confidence interval, 84.3–89.7), specificity 57.7% (47.3–67.7), and AUROC 86.5% (83.5–89.6). The modified G8 had corresponding values of 89.2% (86.5–91.5), 79.0% (69.4–86.6), and 91.6% (89.3; 93.9), with higher AUROC and stable properties in internal (ELCAPA) and external validation (ONCODAGE) compared to the original G8. For the robustness analysis, both tools had AUROCs  $\geq 0.80$  for all definitions tested. Comparing the two instruments, AUROC were significantly higher in favor of the modified G8 to predict 4 out of 6 definitions. Both abnormal scores were independently associated with overall 1-year survival: adjusted hazard ratio [aHR]=4.3[G8]/4.9[modified G8] and 3-years survival: aHR=2.9/2.6; all  $p$ -values $<0.0001$ . Associations persisted after stratifying by metastatic status and in most cancer sites. For both tools, classes of increasing risk showed a graded relationship with mortality.

**Conclusions:** A modified G8 screening tool exhibited better diagnostic performance with better homogeneity across tumor sites and required only six items. These features were confirmed in a large multicenter prospective cohort of older cancer patients, demonstrating the value of the modified tool to identify patients who might benefit from a GA. Our findings demonstrate the robustness of the original and modified G8 to modifications of the reference gold standard, with evidence of a better diagnostic performance of the modified G8 for detecting a variety of health profiles evocative of frailty. Both scores finally demonstrated a strong prognostic value for overall survival, regardless of metastatic status or tumor site. Our findings strengthen the clinical utility of these instruments in the geriatric oncology setting.

**Funding:**

This thesis was realized with a special financial grant from the French National Cancer League (*Ligue Nationale Contre le Cancer*).

The ELCAPA cohort study, data source used in this thesis, was additionally supported by a grant (RINC4) from the French National Cancer Institute (*Institut National du Cancer*, INCa) and *Canceropôle Ile-de-France*, to the Mondor Clinical Research Unit (*URC Henri-Mondor Hospital*).

**Affiliation:**

Clinical Epidemiology and Ageing : Gériatrie, Soins Primaires et Santé Publique (CEpiA) - EA 7376, Université Paris-Est, UPEC.

Address: Service de santé publique, CHU Henri Mondor, 51 Avenue du Maréchal de Lattre de Tassigny, F-94010 Créteil Cedex.



*A mi familia querida,  
que aunque poco han podido verme en los últimos años,  
me apoyan siempre en lo que emprendo y se alegran por mí,*

*a Irving, mi gran compañero, amigo y esposo, por su infinito amor,  
y a Irving Raphael, mi hijo amado que ha hecho que mi vida sea más plena.*



## REMERCIEMENTS (ACKNOWLEDGMENTS)

---

*La réalisation de cette thèse de doctorat a été rendue possible grâce au financement par la Ligue Contre le Cancer du comité du Val-de-Marne. Je tiens à les remercier pour leur confiance et le soutien qu'ils m'ont apportés durant ces trois ans.*

### **A mes directeurs de thèse**

*Tout d'abord, je souhaite exprimer ma profonde gratitude à mes directeurs de thèse, le Pr. Sylvie Bastuji-Garin et le Pr. Etienne Audureau, pour toute leur expérience et leur expertise dont ils m'ont fait profiter.*

*Sylvie, un grand merci pour votre accueil chaleureux au sein de votre équipe en m'accordant votre confiance et votre soutien, qui m'ont permis de mener à bien cette thèse dans d'excellentes conditions. Merci également pour vos précieuses remarques scientifiques et votre disponibilité bienveillante.*

*Etienne, tes qualités scientifiques exceptionnelles, tes qualités humaines et ta gentillesse extraordinaire m'ont permis d'aboutir à la réalisation de ce travail de recherche. Je te suis très reconnaissante pour m'avoir guidée, encouragée, et soutenue, pour tes précieux conseils méthodologiques et ta disponibilité tout au long de cette thèse, mais aussi lors de mon stage de Master 2. J'ai beaucoup appris à tes côtés et je t'adresse ma gratitude.*

### **Aux membres du jury**

*Au Professeur Olivier Guerin et à Carine Bellera pour m'avoir fait l'honneur d'accepter le rôle de rapporteurs de cette thèse et pour leurs remarques pertinentes. Trouvez ici le témoignage de ma reconnaissance.*

*Au Professeur Jean-Luc Novella pour m'avoir fait l'honneur de présider mon jury de thèse. Trouvez ici le témoignage de ma respectueuse considération.*

*Au Professeur Lydia Guittet et à Marie Herr pour m'avoir fait l'honneur d'accepter de juger ce travail. Soyez assurées de ma gratitude.*

## **A l'Ecole Doctorale**

*Mes remerciements vont également à l'Ecole Doctorale de Santé Publique Paris-Sud, à son directeur, Jean Bouyer, et à Audrey Bourgeois, pour leur disponibilité et l'attention qu'ils accordent à chaque doctorant.*

## **Aux membres du laboratoire d'accueil**

*Je remercie le Pr. Elena Paillaud, le Pr. Florence Canoui-Poitrine et le Dr. Philippe Caillet de m'avoir donné l'opportunité de travailler sur les données de la cohorte ELCAPA.*

*A Florence Canoui-Poitrine, Elena Paillaud, Christophe Tournigand, Marie Laurent, Evelyne Liuu, Emilie Ferrat, Rima Ibrahim, Virginie Fossey-Diaz, Stéphane Culine, Jean-Léon Lagrange, Mylène Allain, Anne Chahwakilian, et Pascaline Boudou-Rouquette, pour leur contribution dans les études.*

*A mes collègues de bureau : Nadia, pour tes encouragements, ta gentillesse, ton amitié et tes bons conseils lors des répétitions pour me préparer aux différentes présentations de mes travaux ; Mylène, pour ta bonne humeur et ton aide précieuse de data management ; Richard, pour m'avoir aidée lorsque j'en ai eu besoin ; Florent, pour toutes les discussions partagées et ta musique ; Marie, j'ai mieux dormi grâce à toi, tu sais de quoi je parle !, mais aussi à Nadia et Richard pour avoir été là et m'avoir soutenue ; Cédric, Mathilde, François, Emmanuelle, Lauriane, Benjamin, Camille, Thomas, Aline, Marjolaine, Pierre-André et tous ceux que j'oublie pour votre gentillesse. Merci à tous pour tous les bons moments passés.*

*Je tiens à exprimer mes plus sincères remerciements à Florence Canoui-Poitrine et Anne Puel. Florence et Anne, vous étiez présentes dans les moments les plus difficiles lors des problèmes de santé de mon fils et vous m'avez beaucoup aidée. Merci de m'avoir écoutée lorsque j'en avais besoin. Merci aussi Etienne pour ta compréhension et ton soutien.*

## **A mes amies doctorantes**

*Je n'oublie évidemment pas mes amies doctorantes, Soumaya et Manik : même si nos parcours ont été parfois difficiles et différents, que nous ne nous sommes pas vues autant que nous l'aurions souhaité, vos encouragements et votre amitié ont été sans faille et je vous remercie.*

## **A mes proches**

*A Marina, Philippe, Nathalie, Diana et Erick, merci pour tous les moments extraordinaires, pour nous avoir accompagnés avec autant d'attention depuis notre arrivée; avec vous nous avons eu le Mexique en France. Je vous adresse mes sentiments les plus sincères.*

*A Carole, Thierry, Lena et Theo, je vous remercie beaucoup pour votre amitié et votre gentillesse. Partager des moments avec vous est toujours un vrai plaisir. Merci aussi infiniment pour votre accompagnement depuis notre arrivée en France.*

*A Arnaud, Anne Gaël, Patricia et Dominique, je voudrais vous adresser mes remerciements pour votre aide précieuse et pour vos conseils sur la vie en France, cela nous a beaucoup aidés, mais aussi pour votre amitié et générosité ainsi que pour les moments de joie à vos côtés avec bien sûr Lucas, Louis et Emma.*

*A Christophe et Jeannick pour leur amitié et leur soutien malgré la distance qui nous sépare. Vous nous manquez !*

*Avec vous tous j'ai l'impression de me sentir chez moi.*

*A mi familia querida, que a pesar de la enorme distancia que nos separa, siempre están ahí, dándome su cariño y su apoyo. Mamita querida, papá adorado, gracias por su gran amor, su dulzura, y todo lo que me han enseñado; gracias por compartir mi alegría y celebrar conmigo este y todos los avances en mi carrera. Me siento muy afortunada de tenerlos, son un ejemplo para mí. Gracias a los dos por su confianza. Gracias por el hecho de haber comprendido que la distancia era necesaria aunque fuera por un largo período de tiempo.*

*A mis queridos suegros porque aún desde la distancia he recibido su apoyo y su cariño.*

*A mi pequeño Irving Raphael, por tu sonrisa y tu mirada que me dan la fuerza de pasar todos los obstáculos. Eres el mayor Tesoro que la Vida me ha dado.*

*A mi esposo. ¿Cómo plasmar en solo unas líneas lo que tú has significado para mí en este camino? Gracias por tu paciencia, por tu apoyo incondicional, por tu continua motivación, y por creer en mí. Sin ti a mi lado esto no hubiera sido posible. Eres mi sostén, mi amigo, mi aliado en esta aventura y una fuente constante de estímulo. « Inito » sabrá algún día lo maravilloso que eres. Gracias a los dos por rodearme de tanto amor! Ustedes son mi impulso para mejorar día a día con el objetivo de poder brindarles lo mejor. Los amo.*

*A Dios por todo.*



# CONTENTS

REMERCIEMENTS (ACKNOWLEDGMENTS).....	ix
<b>CONTENTS</b> .....	xiii
LIST OF TABLES.....	xvii
LIST OF SUPPLEMENTAL TABLES.....	xviii
LIST OF FIGURES.....	xix
LIST OF SUPPLEMENTAL FIGURES.....	xix
ACRONYMS AND ABBREVIATIONS.....	xx
SCIENTIFIC VALORISATION.....	xxii
RESUME SUBSTANTIEL EN FRANÇAIS.....	xxiv
CHAPTER 1 Introduction.....	1
1.1 Cancer in the older population: demographic issue.....	2
1.2 Heterogeneity of the older population and the concept of frailty.....	4
1.3 Multidimensional Geriatric Assessment in older cancer patients.....	7
1.3.1 Impact of geriatric interventions on outcomes.....	9
1.3.2 Practical limitations of geriatric assessment in the oncology setting.....	10
1.3.3 Frailty classifications based on GA parameters.....	10
1.4 Screening tools for multidimensional health problems.....	14
1.4.1 What does the term “screening tool” stand for?.....	14
1.4.2 Screening tools in geriatric oncology.....	15
1.5 Research objectives.....	24
1.5.1 Primary objective.....	24
1.5.2 Secondary objectives.....	24
CHAPTER 2 General methodology.....	25
2.1 Study design.....	26
2.1.1 The ELCAPA Prospective Cohort Study.....	26
2.1.2 The ONCODAGE Prospective Multicenter Cohort Study.....	28
2.2 Participants.....	30
2.3 Data collection.....	32
2.4 Data Analysis.....	34
CHAPTER 3 Development of the screening tool modified G8.....	35
3.1 Introduction.....	36
3.2 Methods.....	37
3.2.1 Study population.....	37

3.2.2 GA Reference Procedure .....	37
3.2.3 Candidate Items.....	37
3.2.4 Statistical Analysis.....	38
3.3 Results .....	40
3.3.1 Study population.....	40
3.3.2 Univariate Analysis .....	41
3.3.3 Multiple Correspondence Analysis .....	45
3.3.4 Multivariate Model .....	48
3.3.5 Diagnostic Performance of the Original versus Modified G8.....	48
3.3.6 Internal and Temporal Validation and Sensitivity Analyses.....	49
3.4 Discussion.....	54
3.4.1 Summary of findings.....	54
3.4.2 Comparison with other reviews.....	55
3.4.3 Strengths and limitations.....	56
3.5 Conclusion.....	57
CHAPTER 4 External validation.....	59
4.1 Introduction.....	60
4.2 Methods.....	62
4.2.1 Study population.....	62
4.2.2 Reference standard .....	62
4.2.3 Screening tools.....	62
4.2.4 Statistical analysis .....	62
4.3 Results .....	65
4.3.1 Study population.....	65
4.3.2 Investigating relatedness of development and validation samples .....	67
4.3.3 Assessment of the model's performance in the validation study .....	68
4.3.4 Interpretation of model validation results and updating.....	71
4.3.5 AUROC comparisons.....	72
4.4 Discussion.....	74
4.4.1 Summary of findings.....	74
4.4.2 Strengths and limitations.....	76
4.5 Conclusion.....	78
CHAPTER 5 Gold Standard definitions.....	79
5.1 Introduction.....	80

5.2 Methods.....	81
5.2.1 Study population.....	81
5.2.2 Reference standard definitions.....	81
5.2.3 Screening tools.....	84
5.2.4 Statistical analysis.....	85
5.3 Results.....	87
5.3.1 Study population.....	87
5.3.2 Prevalence of frailty by reference standard.....	89
5.3.3 Predictive performance of screening tools by reference standard.....	90
5.3.4 Sensitivity analyses.....	95
5.4 Discussion.....	95
5.4.1 Summary of findings.....	95
5.4.2 Comparison with other reviews.....	97
5.4.3 Strengths and limitations.....	97
5.5 Conclusion.....	98
CHAPTER 6 Prognostic value of the original and modified G8 tools.....	99
6.1 Introduction.....	100
6.2 Methods.....	101
6.2.1 Study population.....	101
6.2.2 Endpoints.....	101
6.2.3 Screening tools.....	101
6.2.4 Statistical Analysis.....	102
6.3 Results.....	104
6.3.1. Study population.....	104
6.3.2. Survival analyses in the whole study population.....	105
6.3.3. Survival analyses by metastatic status and cancer sites.....	106
6.4 Discussion.....	113
6.4.1 Summary of findings.....	113
6.4.2 Comparison with other reviews.....	113
6.4.3 Strengths and limitations.....	118
6.5 Conclusion.....	119
CHAPTER 7 Discussion.....	121
7.1 The modified G8 compared to other screening tools.....	122
7.1.1 Conceptual design.....	122

7.1.2 Predictive value on outcomes other than GA and mortality .....	128
7.1.3 Direct comparisons between instruments.....	128
7.2 Implications for practice.....	129
7.2.1 Two-step approach .....	129
7.2.2 Modified G8 as part of a 'Geriatric Mini Data-Set' .....	131
7.2.3 Modified G8 as a prognostic tool .....	132
7.3 Strengths and limitations .....	133
CONCLUSIONS AND PERSPECTIVES FOR FURTHER REASERCH.....	135
REFERENCES.....	137
Annex 1. Studies comparing screening tools with Geriatric Assessment in older patients with cancer.....	149
Annex 2. Geriatric Assessment Tools.....	153
Annex 3. Comparisons of baseline characteristics of included and non-included patients in each analysis.....	159
Annex 4. Univariate analysis according to cancer site.....	165
Annex 5. Logistic regression analysis of candidate items predicting impairment of the geriatric assessment.....	171
Annex 6. Multiple Correspondence Analysis .....	173
Annex 7. Complementary results from Gold Standard analysis.....	177
Annex 8. Complementary results from Survival analysis.....	181

## LIST OF TABLES

---

Table 1.1. Domains of geriatric assessment and most frequently used tools .....	8
Table 1.2. Balducci criteria for frailty.....	12
Table 1.3. Factors to be considered for any screening tool. ....	15
Table 1.4. Domains and items evaluated by screening tools to identify older patients with cancer in need of GA .....	17
Table 1.5. Diagnostic performance of screening tools to identify older patients with cancer in need of geriatric assessment.....	21
Table 1.6. G8 screening tool .....	22
Table 3.1. Patient characteristics in the ELCAPA-07 cohort study .....	42
Table 3.2. Univariate analysis of the ability of candidate items to predict impairment of the geriatric assessment (training set; N=729) .....	43
Table 3.3. Final multivariate logistic model for predicting impairment of the geriatric assessment (training set; N=729) .....	50
Table 3.4. Final modified G8: the ELCAPA-07 cohort study .....	51
Table 4.1. Updating methods for predictions models .....	61
Table 4.2. Distribution of patient's characteristics, outcome and predictors .....	66
Table 4.3. Coefficients of the logistic regression model (modified G8 score items) for predicting impairment of GA (development and validation sets).....	67
Table 4.4. Performance of the modified G8 in development and validation datasets.....	69
Table 4.5. Observed vs. Predicted probabilities of an abnormal GA, by decile of risk .....	71
Table 4.6. Comparisons of AUROC in main tumor sites: modified G8 in de validation set vs. development set; original vs. modified G8 in the validation set .....	73
Table 5.1 Frailty classifications approaches (fit vs. "unfit" patients) and indicators.....	84
Table 5.2 Variables and definitions used to classify patients in the latent class typology .....	85
Table 5.3. Patient characteristics.....	88
Table 5.4. Geriatric interventions for overall patient management .....	89
Table 5.5. Associations between different proposals of reference standard and both screening tools (original G8 and modified G8), N=1136.....	92
Table 5.6. Diagnostic performances according to different reference standards (original vs. modified G8).....	94
Table 6.1. Comparison of the prognostic value for overall 1- and 3-years survival of the screening tools G8 and modified G8: unadjusted and adjusted Cox proportional hazards models .....	109
Table 6.2. Comparison of the prognostic value for overall 1- and 3-years survival of the screening tools G8 and modified G8, stratified by metastatic status: unadjusted and adjusted Cox proportional hazards models .....	111

## LIST OF SUPPLEMENTAL TABLES

---

Table A1.1. Diagnostic performances of screening tools against geriatric assessment in older patients with cancer. ....	150
Table A3.1. Patient characteristics comparisons between included and non-included patients in the screening tool development analysis.....	160
Table A3.2. Patient characteristics comparisons between included and non-included patients in the external validation analysis.....	161
Table A3.3. Patient characteristics comparisons between included and non-included patients in the Gold standard definitions analysis.....	162
Table A3.4. Patient characteristics comparisons between included and non-included patients in the survival analysis. ....	163
Table A4.1. Univariate analysis of the association between G8 items and an abnormal Geriatric Assessment, by cancer site.....	166
Table A4.2. Univariate analysis of the association between new candidate items and an abnormal Geriatric Assessment, by cancer site.....	168
Table A5.1. Logistic regression of items to predict impairment of the GA: crude odds ratios, 95%CI and p-value.....	171
Table A6.1. Quality and contribution of categories in the composition of the first two dimensions from multiple correspondence analysis.....	175
Table A7. 1. Associations between indicators and latent class profiles .....	178
Table A7.2. Diagnostic performances according to different reference standards (original vs. modified G8) with available data on the definition tested and both scores.....	179
Table A8.1. Prognostic value of G8: results according to study population with respect to G8/modified G8 availability.....	182
Table A8.2. Prognostic values for overall 1- and 3-years mortality of the ECOG-PS and the TUG: Cox proportional hazards models .....	183
Table A8.3. Prognostic value for overall 1- and 3-years survival by tumor site, using scores (G8 and modified G8) in classes of increasing risk: adjusted Cox proportional- hazards models ...	184
Table A8.4. Harell's C-index and Gönen and Heller's K concordance statistic for G8 and modified G8 by tumor site (overall 1-year survival analysis).....	185
Table A8.5. Harell's C-index and Gönen and Heller's K concordance statistic for G8 and modified G8 by tumor site (overall 3-years survival analysis) .....	186
Table A8.6. Patient characteristics comparisons between patients with and without digestive cancers.....	187

## LIST OF FIGURES

---

Figure 1.1. Share of the population aged 65 and older .....	2
Figure 1.2. Decision tree for treating patients with prostate cancer .....	12
Figure 1.3. Decision tree to determine patient health status .....	13
Figure 2.1. ELCAPA study centers (February 2017) .....	27
Figure 2.2. ONCODAGE study centers (March 2010) .....	28
Figure 2.3. Flow chart of participants included in each study .....	31
Figure 3.1. Receiver-operating characteristic (ROC) curve of number of medications / day for predicting impairment of the geriatric assessment .....	44
Figure 3.2. Multiple correspondence analysis of candidate items .....	46
Figure 3.3. ROC curves for predicting impairment of the geriatric assessment, used as the reference test: original versus modified G8 questionnaire (training set; n = 729) .....	52
Figure 3.4. Area under the ROC curve by cancer site: original versus modified G8 .....	53
Figure 4.1. Calibration plots of the modified G8 applied in the independent cohort (ONCODAGE) .....	70
Figure 4.2. ROC curves for predicting impairment of the geriatric assessment, used as the reference test: original versus modified G8 questionnaire (external validation set; n = 1304) ..	73
Figure 5.1. Receiver-operating characteristic (ROC) curves for predicting different reference standards: original vs. modified G8 questionnaire: .....	93
Figure 6.1. Kaplan-Meier curves of overall 3-years survival for G8 (A) and modified G8 (B) scores .....	108
Figure 6.2. Kaplan-Meier curves of overall 3-years survival for G8 (A) and modified G8 (B) scores, stratified by metastatic status .....	110
Figure 6.3. Prognostic value for overall A) 1- and B) 3-years survival of G8 and modified G8 scores according to tumor sites: unadjusted and adjusted Cox proportional-hazards models ..	112

## LIST OF SUPPLEMENTAL FIGURES

---

Figure A6.1. Scree plot of principal inertias after MCA .....	174
Figure A6.2. Contributions of variables to the Dimension 1. ....	176
Figure A6.3. Contribution of variables to the Dimension 2. ....	176
Figure A8.1. Kaplan-Meier curves of overall 3-years survival for the 6 main tumor sites .....	188

## ACRONYMS AND ABBREVIATIONS

---

**ADL:** Activities of Daily Living

**aCGA:** Abbreviated Comprehensive Geriatric Assessment

**aHR:** adjusted Hazard Ratio

**IADL:** Instrumental Activities of Daily Living

**AUROC:** aire sous la courbe ROC / area under the ROC curve

**BFI:** Brief Fatigue Inventory

**BMI:** Body Mass Index

**CAAF:** Complete arrhythmia with atrial fibrillation

**CES-D:** Center for Epidemiologic Studies Depression Scale

**CHD:** Coronary heart disease

**CHS:** Cardiovascular Health Study

**CI:** confidence interval

**CIRS-G:** Cumulative Illness Rating Scale-Geriatrics

**DSM-IV:** Diagnostic and Statistical Manual of Mental Disorders

**ECOG-PS:** Eastern Cooperative Oncology Group Performance status

**ED:** emergency department

**EG:** Evaluation Gériatrique

**ELCAPA:** Elderly Cancer Patients

**EORTC:** European Organization for Research of Cancer

**FI:** Frailty index of deficit accumulation of Rockwood et al.

**fTRST:** Flemish version of the Triage Risk Screening Tool

**G8:** Geriatric-8

**GA:** Geriatric Assessment

**GDS:** Geriatric Depression Scale;

**GFI:** Groningen Frailty Indicator

**HL:** Hosmer-Lemeshow test

**HR:** Hazard Ratio

**IC 95%:** Confidence interval 95%

**INCa:** French National Cancer Institute

**IQR:** interquartile range

**ISAR:** Identification of Seniors At Risk

**KPS:** Karnofsky Performance Status  
**LP:** Linear predictor  
**LR+:** Positive Likelihood Ratio  
**LR-:** Negative Likelihood Ratio  
**MAPE:** Erreur absolue moyenne / Mean Absolute Prediction Error  
**MMSE:** Mini Mental State Examination  
**MNA:** Mini Nutritional Assessment  
**MNA-SF:** Mini Nutritional Assessment short form  
**MOB-T:** Mobility Tiredness Test  
**MOS:** Medical Outcomes Study  
**NPV:** Negative Predictive Value  
**OR:** Odds ratio  
**OGS:** Onco-geriatric screen  
**PPV:** Positive Predictive Value  
**PS:** Performance Status  
**QLQ C-30:** Quality-of-Life Questionnaire Core Questionnaire 30 items  
**ROC:** Receiver-operating curve  
**RPA:** Recursive partitioning analysis  
**SAOP2:** Senior Adult Oncology Program 2 screening  
**SD:** standard deviation  
**SOF:** Study of Osteoporotic Fractures Index  
**SIOG:** International Society of Geriatric Oncology  
**TUG:** Timed Up-and-Go test  
**UCOG:** Unité de Coordination en Oncogériatrie  
**VES-13:** Vulnerable Elders Survey

### Published articles

- [1] **Martinez-Tapia C**, Canoui-Poitrine F, Bastuji-Garin S, Soubeyran P, Mathoulin-Pelissier S, Tournigand C, Paillaud E, Laurent M, Audureau E, and on behalf of the ELCAPA Study Group. Optimizing the G8 screening tool for older patients with cancer: Diagnostic performance and validation of a six-item version. *The Oncologist*. 2016 Feb; 21(2):188–195.
- [2] **Martinez-Tapia C**, Paillaud E, Liuu E, Tournigand C, Ibrahim R, Fossey-Diaz V, Culine S, Canoui-Poitrine F, Audureau E, and on behalf of the ELCAPA Study Group. Prognostic value of the G8 and modified G8 screening tools for multidimensional health problems in older patients with cancer. *European Journal of Cancer*. 2017 Jul; 83:211 – 219.

### Article submitted

- [3] **Martinez-Tapia C**, Laurent M, Ferrat E, Paillaud E, Lagrange JL, Allain M, Chahwakilian A, Boudou-Rouquette P, Bastuji-Garin S, Audureau E, and on behalf of the ELCAPA Study Group. Screening for multidimensional health problems in older patients with cancer: effect of varying gold standard definitions on the diagnostic performance of the G8 and Modified G8 screening tools. (Submitted to British Journal of Cancer)

### Article in progress

- [4] **Martinez-Tapia C**, Audureau E. External validation of the 6-Item Modified G8 for screening frailty in a large multicentric prospective cohort of older patients with cancer.

## Oral presentations in congresses and seminars

**Martinez-Tapia C**, Canoui-Poitrine F, Bastuji-Garin S, Soubeyran P, Mathoulin-Pelissier S, Paillaud E, Laurent M, Audureau E. Modification of the G8 screening tool for frailty in elderly patients with cancer: The Elcapa-07 cohort study. *Journal of Geriatric Oncology*. 2014; 5(2):S15 (Abstract 008). (SIOG congress, Lisbon, Portugal, 23/10/2014).

**Martinez-Tapia C**, Canoui-Poitrine F, Bastuji-Garin S, Soubeyran P, Mathoulin-Pelissier S, Tournigand C, Paillaud E, Laurent M, Audureau E. Optimisation de l'outil de dépistage G8 chez les patients âgés atteints de cancer : l'étude de cohorte ELCAPA-07. Journées 2015 du groupe de travail Oncogériatrie du Cancéropôle Ile-de-France (« Oncogériatrie : de la recherche à la pratique » ; Paris 19/06/2015)

**Martinez-Tapia C**, Canoui-Poitrine F, Bastuji-Garin S, Soubeyran P, Mathoulin-Pelissier S, Tournigand C, Paillaud E, Laurent M, Audureau E. Optimisation de l'outil de dépistage G8 chez les patients âgés atteints de cancer : l'étude de cohorte ELCAPA-07. *L'année gériatrique*. 2016; 30(2):16 (Abstract 04-1). 36èmes journées annuelles de la Société Française de Gériatrie et Gériatrie (SFGG 2016 – Paris, France)

## Posters

**Martinez-Tapia C**, Canoui-Poitrine F, Bastuji-Garin S, Soubeyran P, Mathoulin-Pelissier S, Tournigand C, Paillaud E, Laurent M, Audureau E. Optimisation de l'outil de dépistage G8 chez les patients âgés atteints de cancer : l'étude de cohorte ELCAPA-07. 12èmes Journées de la Société Francophone d'Onco-Gériatrie (SoFOG 2016 – Montpellier, France).

**Martinez-Tapia C**, Laurent M, Paillaud E, Lagrange JL, Allain M, Chahwakilian A, Boudou-Rouquette P, Bastuji-Garin S, Audureau E. Screening for multidimensional health problems in older patients with cancer: effect of varying gold standard definitions on the diagnostic performance of the G8 and Modified G8 screening tools. *Journal of Geriatric Oncology*. 2016; 7/6(1):S91 (Abstract P092) (Congrès de la SIOG, Milan, 2016).

### **Introduction**

L'évolution de la population française a été caractérisée au cours des dernières décennies par un processus de vieillissement continu, consécutif à la baisse de la fécondité et au recul de l'espérance de vie constaté dans la plupart des pays industrialisés. Au 1er janvier 2017, la France compte 66,9 millions d'habitants, dont 19,2% ont 65 ans ou plus, soit une augmentation de 3% en dix ans [1]. Du fait du vieillissement de la population et de l'augmentation du risque de développer un cancer avec l'âge, plus de 60% des nouveaux cas de cancer sont actuellement diagnostiqués chez les personnes âgées de plus de 65 ans [2,3], ce qui constitue un enjeu de santé publique pour la prise en charge optimale de ces patients.

### **Le concept de fragilité en oncogériatrie**

De nombreux travaux ont fait le constat d'une hétérogénéité majeure dans les capacités d'adaptation des sujets âgés à la maladie cancéreuse et la tolérance aux traitements entrepris [4,5]. Un large domaine de recherche s'est ainsi développé autour des concepts de diminution des réserves fonctionnelles, de vulnérabilité ou « fragilité ». La caractéristique essentielle de la fragilité est une sensibilité accrue aux facteurs de stress, liée à des déficiences dans différents systèmes et à une diminution des réserves physiologiques. Les sources de vulnérabilité les mieux connues incluent les maladies chroniques (comorbidité), la dépendance fonctionnelle, l'état nutritionnel, la détérioration cognitive et la dépression [6].

S'il n'existe pas de consensus quant au périmètre de ces concepts – notamment vis-à-vis de la multimorbidité –, il est néanmoins admis que ces éléments de fragilité peuvent avoir un impact sur la tolérance de la thérapie anticancéreuse et ainsi modifier le rapport bénéfice-risque du traitement [7]. Afin de garantir un traitement optimal et adapté au patient âgé atteint de cancer, il est donc primordial de reconnaître l'hétérogénéité de cette population et repérer les patients les plus fragiles.

## L'évaluation gériatrique

Dans ce contexte, l'Evaluation Gériatrique (EG) a été développée par les gériatres dans l'objectif de caractériser ces éléments de fragilité et, *in fine*, aider à la décision médicale chez les patients âgés atteints de cancer [8]. Même s'il n'existe pas de définition univoque et définitive des composantes de l'EG, elle consiste en une évaluation multidimensionnelle de l'état de santé du patient comportant l'évaluation à l'aide de questionnaires ou tests validés du statut fonctionnel, de l'autonomie, de la mobilité, de l'état nutritionnel, des comorbidités et des troubles cognitifs et psychologiques. Une méta-analyse a montré que l'intégration de l'EG en pratique gériatrique était associée à une survie prolongée, une amélioration du statut fonctionnel et une diminution du taux d'institutionnalisation chez les patients âgés non atteints de cancer [9]. Chez les patients âgés atteints de cancer, l'EG constitue un outil gériatrique utile pour distinguer les patients susceptibles de bénéficier d'un traitement standard des patients à risque élevé de complications et/ou trop vulnérables pour recevoir un traitement agressif [10,11], pour décider de la mise en place des soins de support pouvant améliorer la tolérance au traitement [6,12], et en termes d'aide à la décision thérapeutique et d'efficacité globale (survie et qualité de vie) [12-14].

Le *National Comprehensive Cancer Network* et la Société Internationale d'Oncogériatrie (SIOG) recommandent l'utilisation d'une EG multidimensionnelle avant l'élaboration d'un programme personnalisé de soins pour tous les patients cancéreux de 70 ans et plus, afin d'identifier d'éventuels problèmes de santé ou déficiences (e.g. facteurs de fragilité) et ainsi optimiser la stratégie thérapeutique [10,15,16].

## **Les outils de dépistage de la fragilité**

Si l'EG est actuellement considérée comme la procédure de référence pour identifier les patients âgés vulnérables atteints de cancer, elle nécessite cependant une expertise pour l'ensemble des domaines explorés et est de fait particulièrement chronophage pour le gériatre en pratique clinique courante (entre 1h30 et 2 heures par patient). C'est pourquoi une partie de la recherche actuelle en oncogériatrie se concentre sur la construction d'outils de dépistage pratiques et simples, visant à discriminer en amont de l'EG les patients vulnérables pour lesquels une évaluation complète pourrait permettre une adaptation du plan de traitement oncologique.

Ces outils devraient idéalement combiner une bonne sensibilité afin d'assurer une identification correcte des patients vulnérables, mais également une spécificité suffisante pour limiter le nombre de faux positifs et ainsi préserver le temps consacré à l'EG dans un contexte de nombre restreint de gériatres disponibles.

## **Limites des instruments actuels**

Dans une revue systématique récente portant sur les instruments de dépistage de la fragilité actuellement disponibles et utilisables chez les patients âgés atteints de cancer, Hamaker et al ont montré que la performance diagnostique de ces instruments de

mesure restait très insuffisante vis-à-vis de l'EG [17]. Parmi ces outils, le G8 a été développé spécifiquement pour les sujets âgés atteints de cancer et bénéficie d'une sensibilité élevée pour la détection des patients chez qui une EG complète serait nécessaire. Néanmoins, sa spécificité reste faible pour la pratique clinique.

Par ailleurs, une grande variabilité est constatée quant au gold standard retenu pour définir la « fragilité », limitant de fait la comparabilité entre instruments, et dans un contexte où les résultats d'analyses de sensibilité permettant d'apprécier la robustesse des outils au changement de gold standard ne sont qu'exceptionnellement publiés. Enfin, le choix des items retenus pour la plupart des outils existants repose presque exclusivement sur l'opinion d'experts en oncogériatrie, approche indispensable pour la pertinence de l'outil, mais sans qu'un développement statistique complémentaire ait permis d'en optimiser les performances. A ce titre, l'utilisation de techniques psychométriques spécifiques - analyse de la redondance des items, analyses multidimensionnelles, etc. - pourrait permettre une amélioration substantielle des performances de ces instruments.

---

## Synthèse

- L'évaluation gériatrique standardisée est recommandée chez les patients âgés atteints de cancer mais reste particulièrement **chronophage** compte tenu des multiples composantes et échelles à réaliser pour son évaluation.
- Il existe un réel besoin pour une **échelle de dépistage de la fragilité** combinant *rapidité* de réalisation, *performance métrologique* et *pertinence clinique*. Néanmoins :
  - 1) Aucune échelle de dépistage actuellement utilisée en oncologie gériatrique ne combine une **sensibilité et spécificité adéquate** pour la détection de la fragilité, quelle que soit sa définition.
  - 2) Ces échelles ont le plus souvent été développées sur la base exclusive d'opinions d'experts **sans développement statistique psychométrique** spécifique.
  - 3) Une variabilité importante est constatée quant à l'opérationnalisation du concept de « fragilité » mais il n'existe que peu de données permettant d'apprécier la **robustesse des outils de dépistage au changement de gold standard**.

## **Objectif général de la thèse**

L'objectif principal de cette recherche était ainsi de développer et valider un nouvel outil de dépistage performant et utilisable en oncologie gériatrique qui permettrait d'améliorer l'identification précoce des patients avec un profil de risque gériatrique, et chez qui une EG complète pourrait être réalisée afin d'assurer une prise en charge plus adaptée du cancer et des déficiences associées.

Pour réaliser cet objectif, un état des lieux complet des échelles de dépistage de la fragilité existantes a été réalisé, afin de recenser les domaines et items évalués, les définitions utilisées pour caractériser l'état de fragilité à dépister (gold standard) et les méthodes utilisées pour leur validation. Une approche statistique systématique a ensuite été conduite afin de développer un nouvel outil sur la base des items du G8 original les plus discriminants et d'un ensemble d'items supplémentaires cliniquement pertinents. Les propriétés diagnostiques du nouvel outil (le G8 modifié) ont été vérifiées par validation interne sur les données de la cohorte de développement ELCAPA et validation externe à partir des données de la cohorte ONCODAGE.

## **Objectifs secondaires**

Les objectifs secondaires visent à documenter la robustesse, la valeur pronostique et l'utilité du nouvel outil, incluant :

- étudier la robustesse de l'instrument vis-à-vis des changements du gold standard ;
- étudier la validité pronostique de cette nouvelle échelle vis-à-vis de la mortalité à court et à moyen terme.

## Méthodes

### Sources des données

Pour réaliser ce travail de thèse, les données de deux cohortes prospectives de patients âgés atteints de cancer ont été exploitées :

- La cohorte ELCAPA (ELderly CAncer Patients ; N=2952 en 02/2017) est une cohorte ouverte prospective, dynamique, hospitalière multicentrique de l'ensemble des patients âgés de 70 ans ou plus atteints de cancer recrutés consécutivement à partir de Janvier 2007, et adressés par un oncologue, radiothérapeute, chirurgien ou spécialiste d'organe à l'unité de coordination en oncogériatrie (UCOG) pour réalisation d'une EG.
- La cohorte ONCODAGE (N=1435) est une étude de cohorte, prospective, multicentrique portant sur les patients âgés de 70 ans et plus atteints de cancer recrutés consécutivement entre août 2008 et mars 2010 pour la validation de l'outil de dépistage G8 d'origine. Pour ce travail de thèse ont été inclus les patients en 1er temps de prise en charge thérapeutique ou entre deux étapes d'une séquence de traitement de première ligne prédéfinie.

### Analyses statistiques

Pour le développement du nouvel outil, le G8 modifié, une évaluation des performances diagnostiques de chacun des items candidats a été réalisée, suivi par l'optimisation du codage des items, l'étude des corrélations et redondances entre les items, le calcul des pondérations et la construction du score final. Des analyses en correspondances multiples ont été réalisées, ainsi que des analyses multivariées avec

utilisation de techniques d'imputation multiple des données manquantes. Une validation interne de l'outil par techniques de bootstrap a été réalisée ainsi qu'une validation temporelle à partir de nouveaux patients recrutés au sein de la cohorte ELCAPA.

La validation externe de l'outil a été réalisée en examinant la discrimination et calibration du G8 modifié au sein de la cohorte ONCODAGE. La discrimination, capacité du modèle à séparer les patients qui présentent ou non un profil de risque de fragilité, a été évaluée par l'aire sous la courbe ROC (AUROC), la sensibilité, la spécificité, la valeur prédictive positive et négative, et le rapport de vraisemblance positif et négatif. La calibration, visant à comparer les probabilités prédites par le modèle d'avoir une EG anormal et les fréquences réellement observées, a été évaluée par l'erreur absolue moyenne (MAPE) entre les probabilités prédites et observées, ainsi que par le calcul de l'intercept, la pente de calibration et le test de Hosmer-Lemeshow. La performance globale du modèle a été évaluée par le score de Brier.

La stabilité de la performance diagnostique du G8 modifié et du G8 original a été évaluée vis-à-vis d'autres situations évocatrices d'un état de fragilité et cliniquement pertinentes : 1) 2 anomalies à l'EG, 2) la prescription d'au moins une intervention significative par le gériatre, 3) l'existence d'un profil de fragilité tel que défini par une approche en classes latentes et deux autres classifications basées sur l'avis d'experts et utilisés en oncologie gériatrique (Balducci et SIOG). Les AUROC ont été calculées et comparées entre les deux outils par rapport aux six définitions de référence testés. Les seuils optimaux pour chaque outil (en priorisant la sensibilité) ont été déterminés pour estimer les sensibilités et spécificités vis-à-vis de chaque Gold Standard, qui ont été comparées par le test de Chi<sup>2</sup> de McNemar.

Enfin, la survie globale à 1 an et 3 ans a été analysée par les courbes de Kaplan-Meier en comparant chaque groupe par des tests du log-rank. Des modèles univariés de Cox à risques proportionnels ont été réalisés pour les deux outils, puis ajustés sur l'âge, le sexe, le traitement anticancéreux et le statut métastatique du patient. Des analyses stratifiées sur le statut métastatique ont été également effectués. La performance discriminative a été évaluée par l'index C de Harell et la statistique de concordance K de Gönen et Heller. Toutes les analyses ont été menées globalement et par localisation tumorale.

## Résultats

La phase de développement a permis de construire une version modifiée de l'outil de dépistage G8 comprenant six facteurs prédictifs d'une Evaluation Gériatrique (EG) anormale et simples à recueillir par les gériatres : perte de poids au cours des 3 derniers mois, existence de problèmes neuropsychologiques, statut clinique *Eastern Cooperative Oncology Group Performance Status* (ECOG-PS en version simplifiée), état de santé perçu, poly-prescription ( $\geq 6$  médicaments / jour), et existence parmi les antécédents d'une insuffisance cardiaque / maladie coronarienne. Avec un seuil de  $\geq 6 / 35$  points, le G8 modifié a démontré à la fois une sensibilité et spécificité élevées (89,2% et 79,0% respectivement) avec une AUROC de 0,92 (IC 95% : 0,89 à 0,94). Le nouvel outil montrait une bonne homogénéité entre les sites tumoraux, y compris les tumeurs solides et les cancers hématologiques (cohorte ELCAPA, période d'inclusion : Janvier 2007 - Octobre 2012).

La validation externe du score G8 modifié a été réalisée dans un échantillon de 1304 patients issus de la cohorte multicentrique ONCODAGE. Par rapport à la cohorte de développement, la cohorte de validation externe comptait significativement plus de patients atteints de cancer du sein ou du poumon et moins de patients atteints de cancer du tractus digestif haut/foie ou localisations urologiques ( $p < 0,0001$ ). Les patients étaient plus jeunes, avaient moins souvent de métastase(s) et comptaient majoritairement des femmes. Ils avaient un meilleur état général mesuré par l'EGOS-PS ( $p < 0,0001$ ), avec une EG anormale dans 79,4% des cas comparativement à 86,7% dans la cohorte de développement. Le G8 modifié a montré de bonnes performances diagnostiques avec une sensibilité de 82,2% (IC 95% : 79,8 à 84,5), une spécificité de 89,5% (63,3 à 74,6) et une AUROC de 0,85 (0,82 à 0,87). En dehors du premier décile, où le score sous-estimait le risque d'avoir une EG anormale, la calibration du score était bonne après ajustement de l'intercept à la prévalence d'une EG anormale dans la cohorte de validation.

Concernant l'étude de robustesse vis-à-vis du changement de gold standard, l'AUROC était égale ou supérieure à 0,80 pour les deux outils et toutes les définitions testées (cohorte ELCAPA, période d'inclusion : Janvier 2007 – Juin 2015). En comparant les deux instruments, les AUROC's étaient significativement plus élevées en faveur du G8 modifié pour la prédiction de 4 des 6 définitions testées: EG  $\geq 1$  anomalie, (G8 modifiée: 0,93 [IC 95% 0,91 à 0,95] vs. G8 d'origine: 0,90 [0,87 à 0,92];  $p=0,0029$ ), EG  $\geq 2$  anomalies (0,90 [0,88 à 0,92] vs. 0,87 [0,88 à 0,92];  $p=0,0006$ ),  $\geq 1$  intervention significative prescrite (0,85 [0,81 à ,89] vs. 0,81 [0,77 à 0,86];  $p=0,0056$ ) et des patients fragiles selon la classification de Droz (0,88 [0,86 à ,91] vs. 0,83 [0,81 à 0,86];  $p < 0,0001$ ). Ces résultats attestent de la robustesse des deux instruments vis-à-vis du changement

du Gold Standard, avec une meilleure performance diagnostique pour le G8 modifié pour détecter une variété de profils de santé évocateurs d'une fragilité.

Enfin, des résultats comparables entre le G8 original et le G8 modifié ont été trouvés en termes de valeur pronostique vis-à-vis de la mortalité à 1 an (HR ajusté = 4,31 [IC à 95% 2,73 à 6,8];  $p < 0,0001$  pour le G8 original ; HR ajusté = 4,87 [3,1 à 7,64];  $p < 0,0001$  pour le G8 modifié) et à 3 ans (HR ajusté = 2,94 [2,17 à 3,98] ;  $p < 0,0001$  pour le G8 original ; HR ajusté = 2,56 [1,95 à 3,37] ;  $p < 0,0001$  pour le G8 modifié) dans la cohorte ELCAPA (période d'inclusion : Janvier 2007 - Avril 2014). Les deux outils ont montré des fortes associations avec la mortalité à 1 et 3 ans pour les cancers de la prostate, cancers urinaires, du sein et les hémopathies (ex: cancer de la prostate : performance discriminative C-Harrell de 0,84 pour le G8 modifié). Pour les cancers colo-rectaux, seul le G8 modifié était significativement associé avec la mortalité à 1 an (C-Harrell = 0,72). Aucune association significative n'était retrouvée pour les tumeurs de l'appareil digestif haut en considérant les seuils binaires (score normal/anormal), mais les associations pour les deux outils étaient statistiquement significatives pour tous les types de cancer en considérant les scores en classes de risque croissant. Les résultats de ces analyses renforcent la validité et l'utilité de l'instrument G8 modifié en tant qu'outil d'aide pour l'oncologue et le gériatre pour la prise en charge du patient et aide à la décision thérapeutique.

## **Conclusion**

Les travaux accomplis au cours de cette thèse ont permis de développer une version modifiée du G8 bénéficiant d'une amélioration substantielle de la performance diagnostique, globalement et pour toutes les localisations cancéreuses testées, et d'une plus grande parcimonie avec seulement six items à recueillir au lieu de huit pour le G8 original. La bonne performance diagnostique du nouvel outil a été validée dans une cohorte indépendante de patients âgés atteints de cancer et l'outil a démontré sa robustesse pour détecter une variété de profils de santé évocateurs de fragilité et sa valeur pronostique forte pour la mortalité à 1 et 3 ans.

Notre travail illustre l'intérêt de combiner des analyses statistiques approfondies aux jugements d'experts afin d'assurer à la fois discrimination optimale, valeur pronostique et utilité clinique. L'utilisation du G8 et G8 modifié devrait être encouragée en oncologie gériatrique afin d'étudier l'impact sur la prise en charge et l'état de santé des patients âgés de cancer d'une implémentation de ces outils de dépistage en amont et/ou association systématique à l'évaluation gériatrique.

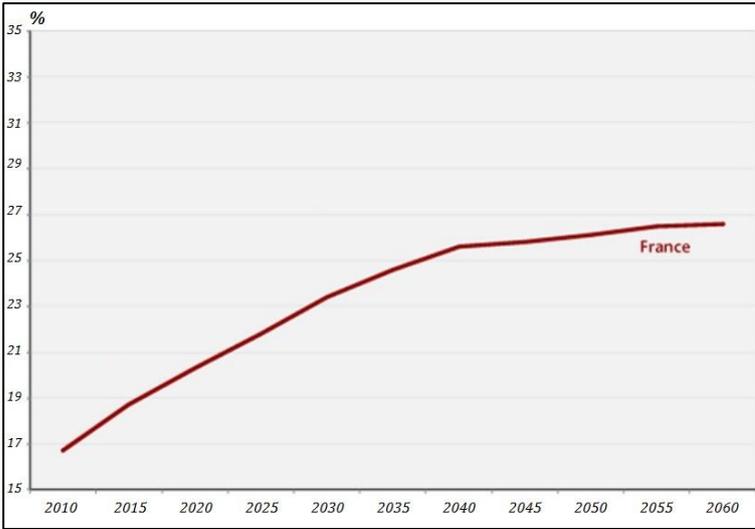
# CHAPTER 1

---

## **Introduction**

### 1.1 Cancer in the older population: demographic issue

The world’s population is ageing rapidly. In 2016, an estimated 636 million people were aged 65 years and older (8%) [4]. By 2050, this number is projected to nearly triple to about 1.5 billion, representing 16% of the world’s population [4,5]. The number of people aged 80 years and older is growing even faster. Projections from the United Nations statistics indicate that the number of persons 80 years and older will more than triple from the current 135 million to 434 million in 2050 [4,6]. In France for example, over the period 2011-2015, the share of the population aged 65 and over increased by 12% [7], and will continue to increase to more than 25% in the next 25 years (**Figure 1.1**).



Source: European Commission, "The 2012 Aging Report"

**Figure 1.1.** Share of the population aged 65 and older

There is reason for concern regarding an anticipated increase in the number of cancer cases, since the majority and growing number of cancers are diagnosed in patients aged 65 years and older [8], with a median age of diagnosis between 65 and 69 years [7,9,10].

In 2012, there were an estimated 6.7 million new cases of cancer in people aged 65 years and older, worldwide: 3.9 million (58%) in males and 2.8 million (42%) in females, and it is predicted that there will be 11.8 million new cancer cases each year by 2030, if recent trends in incidence of major cancers and population growth are seen globally in the future. This is 76% more cases than in 2012. In France, there were 233 943 new cases of cancer in people aged 65 and older in 2015 (58% in males), representing 61% of all cancers diagnosed [2]. Indeed, people aged 65 years and older account for 60% of newly diagnosed cancer cases and about 70% of all cancer deaths [2,3]. In 2015, 112 596 deaths from cancer are estimated in France among people aged 65 and older, accounting for 75% of all cancer deaths [2]. The cancers with the highest mortality in men aged 65 years and older, in descending order were lung, prostate and colorectal cancers, together accounting for almost half (48%) of all cancer deaths in men of this age group. In women aged 65 years and older, these cancers concerned breast, colorectal and lung (43% of all cancer deaths in women of this age group) [2].

Population aging and the increase in cancer incidence with advancing age make the management of older patients with cancer a major public health challenge. This increased number of older patients with cancer will need physicians and healthcare professionals to have special expertise in both oncology and geriatrics and so, better understand the management of this population.

## 1.2 Heterogeneity of the older population and the concept of frailty

The population of older patients with cancer varies widely with respect to multiple aspects of the aging process and health status. This is a very heterogeneous group, ranging from fit patients who are competent, active, with no functional deficits and few or any comorbidities and geriatric syndromes, to frail patients with multiple chronic conditions and geriatric syndromes, who are cognitively impaired and dependent [11]. This heterogeneity contributes to the complexity to determine the optimal cancer treatment for these patients. Generally, it is considered appropriate for fit patients to receive the same treatments used in younger adults, whereas frail patients are at higher risk for toxicity from standard treatment [12]. A large number of studies have already pointed out the heterogeneity in elderly's ability to adapt to life with cancer and to tolerate treatments [13,14]. Identification of those patients at higher risk of poor outcomes and individual treatment tailoring are therefore particularly important to ensure optimal efficacy and to minimize toxicity in this age group [15].

Until recently, there has been no consensus on the definition of frailty for older people, which has been described as a state of increased vulnerability to stressors resulting from decreased physiological reserves in multiple organ systems, leading to increased risk of adverse health outcomes such as functional decline, disability, falls, hospitalization, and death [16,17]. In an effort to reach consensus around frailty's definition, a consensus group consisting of delegates from 6 major international, European, and US geriatric societies recently suggested a definition around the concept of "physical frailty", which was defined as *"a medical syndrome with multiple causes and contributors that is characterized by diminished strength, endurance, and reduced physiologic function that increases an individual's vulnerability for developing increased*

*dependency and/or death*" [18]. The panel identified 3 other major points: 1) physical frailty can be prevented or treated; 2) it can be recognized by validated screening scales; and 3) all persons older than 70 years and people with weight loss of greater than 5% should be screened for frailty. Although most works have focused on the physical domain, others also considered cognition in the definition of frailty. An international consensus group organized by the International Academy on Nutrition and Aging (I.A.N.A) and the International Association of Gerontology and Geriatrics (I.A.G.G) defined "cognitive frailty" as *"an heterogeneous clinical manifestation characterized by the simultaneous presence of both physical frailty and cognitive impairment"* [19]. Two defining criteria were proposed: 1) presence of physical frailty and cognitive impairment; and 2) exclusion of concurrent Alzheimer's disease or other dementias. Other definitions suggested that an integral approach is needed for the concept of frailty, which should incorporate several domains, and is thus based on the integral functioning of the individual [20]. More recently, in order to clarify the definition of frailty for older patients with cancer in the context of oncology trials, the Cancer and Aging Research Group in collaboration with the National Cancer Institute, the National Institute on Aging, and the Alliance for Clinical Trials in Oncology, suggested a geriatric oncology definition of frailty: *"those older individuals who are at higher risk for cancer treatment toxicity because of age associated conditions such as functional losses, cognitive impairment, or physiologic changes"* [21].

Although various definitions of frailty have been proposed and the concept is accepted with broad agreement, how to actually measure this concept remains controversial. There is still no single operational definition of frailty, varying widely according to the conceptual framework. Two major approaches to measure frailty in older people exist in the literature: the frailty phenotype, developed by Fried et al. [22]

and conceptualized around an evidence-based biologic pathway in which aging-related cellular and physiological changes lead to frailty, and the frailty index of deficit accumulation (FI) of Rockwood et al. [23], whose concept presupposes that a global system loses robustness as it develops various illnesses or functional declines. The former is a 5-component tool including unintentional weight loss (more than 10 pounds in the previous year), weakness (grip strength in the lowest quartile as determined by dynamometric measurement), exhaustion (measured by questions about energy levels from a depression survey), slow walking speed (slowest quartile of performance on a timed 15 feet walk), and low levels of physical activity (measured by the Minnesota Leisure Time Activities questionnaire). An older individual is considered frail when at least three of these criteria are present, and robust when they have none. An intermediate level, in which one or two criteria are present, identifies a possible pre-frail stage at high risk of progressing to frailty. In contrast to Fried's criteria, which are mostly based on physical parameters, the FI comprises a checklist of 70 items of clinical conditions and diseases including disability, physical and cognitive impairments, psychosocial risk factors, and geriatric syndromes (e.g. falls, delirium, and urinary incontinence). The index is created by counting the accumulation of deficits that can be cumulatively scored. A higher frailty score indicates a greater degree of frailty.

Both measures have been validated in large population-based aging cohorts through their association with adverse health outcomes [24,25]. However, neither of the two measures was developed specifically for older patients with cancer in whom physiologic stressors may differ from the general geriatric population, limiting their applicability to these patients for whom a multidimensional geriatric assessment (GA) is more appropriate [15].

### 1.3 Multidimensional Geriatric Assessment in older cancer patients

The geriatric assessment was originally developed by geriatricians as a multidisciplinary evaluation of older patients [26]. Since 2005, the International Society of Geriatric Oncology (SIOG) recommends a GA for every cancer patient aged  $\geq 70$  years [27]. Current guidelines from national (the National Comprehensive Cancer Network, United States) and international societies (SIOG) recommend performing a multidimensional GA to assess the profile of each older patient with cancer before setting up an individualized care program [28,29]. GA produces an inventory of health problems, capturing general health status with the assessment of comorbidities, functional, cognitive, social, nutritional, and psychological parameters [26]. It can help detecting geriatric syndromes and unaddressed impairments requiring adequate management and support [27,28,30]. GA can also be valuable in oncology practice for predicting treatment-related toxicity and feasibility, functional decline and overall survival in various tumors and treatment settings [28,31-34]. GA can influence treatment choice and intensity [35] and allow targeted interventions, which can in turn improve quality of life and compliance to therapy [28,36,37].

The GA uses validated tools to assess the different domains comprised in the evaluation. Two systematic reviews by the SIOG GA task force summarized the domains one should evaluate in the geriatric assessment and all available tools related to GA in geriatric oncology [28,38]. **Table 1.1** presents the important domains that a GA should comprise along with the most frequently used questionnaires or instruments for assessing these domains.

**Table 1.1.** Domains of geriatric assessment and most frequently used tools

<b>Domains</b>	<b>Most frequently used tools<sup>a</sup></b>
1. Social status	<ul style="list-style-type: none"> <li>– Living arrangements (i.e. living alone, availability of a caregiver, appropriateness of social environment: 16 studies)</li> <li>– Availability of social support (MOS: 10 studies)</li> </ul>
2. Functional status	<ul style="list-style-type: none"> <li>– ADLs (82 studies): most often measured by the Katz index</li> <li>– IADLs (80 studies): most often measured by the Lawton scale</li> <li>– Performance status (45 studies): most often measured by the ECOG-PS and the Karnofski scale</li> <li>– Gait speed (23 studies): most often measured by the Timed Get Up and Go test.</li> </ul>
3. Comorbidity	<ul style="list-style-type: none"> <li>– Charlson comorbidity index (26 studies)</li> <li>– CIRS(-G) (25 studies)</li> </ul>
4. Cognition	<ul style="list-style-type: none"> <li>– MMSE (54 studies)</li> </ul>
5. Depression	<ul style="list-style-type: none"> <li>– GDS (any version: 48 studies)</li> </ul>
6. Nutrition	<ul style="list-style-type: none"> <li>– Body Mass Index (24 studies)</li> <li>– MNA (23 studies)</li> <li>– Weight loss (7 studies)</li> </ul>
7. Polypharmacy	<ul style="list-style-type: none"> <li>– Total number of prescribed medications (22 studies)</li> </ul>
8. Fall risk assessment	<ul style="list-style-type: none"> <li>– Self-reported falls (26 studies)</li> </ul>
9. Fatigue	<ul style="list-style-type: none"> <li>– 9 studies, each using a different instrument: SF36, MOB-T, EORTC, EORTC QLQ C-30, BFI, 2-items CES-D, Visual Analogue Scale, or tool not specified.</li> </ul>
10. Geriatric syndromes	<ul style="list-style-type: none"> <li>– Dementia: 10 studies</li> <li>– Delirium: 9 studies</li> <li>– Incontinence (fecal and/or urinary): 12 studies</li> <li>– Osteoporosis or spontaneous fractures: 10 studies</li> <li>– Neglect or abuse: 7 studies</li> <li>– Failure to thrive: 5 studies</li> </ul>

Abbreviations: MOS, Medical Outcomes Study; ADL, Activities of Daily Living; IADL, Instrumental ADL; ECOG-PS, Eastern Cooperative Oncology Group; CIRS-G, Cumulative Illness Rating Scale-Geriatrics; MMSE, Mini Mental State Examination; GDS, Geriatric Depression Scale; MNA, Mini Nutritional Assessment; MOB-T, Mobility Tiredness Test; EORTC, European Organization for Research of Cancer; QLQ C-30, Quality-of-Life Questionnaire Core Questionnaire 30 items; BFI, Brief Fatigue Inventory; CES-D, Center for Epidemiologic Studies Depression Scale.

<sup>a</sup> According to 95 studies included in the reviews of Puts et al. [38] and Wildiers et al. [28] (published between January 1, 1996 and March 7, 2013).

A Delphi consensus of geriatric oncology experts was consistent with these two previous reviews [39]. The expert panel concluded that all domains among those evaluated (the first seven domains of Table 1.1) should be included in a GA in order to guide care processes. Consensus was met similarly for the tools that are frequently and appropriately used to assess each domain in clinical practice as part of a GA.

### 1.3.1 Impact of geriatric interventions on outcomes

Studies that assessed the influence of GA in the care of the older patient with non-malignant diseases have been summarized in two meta-analyses. The first included 28 geriatric randomized controlled trials, which demonstrated that GA, if linked to geriatric interventions, reduced early rehospitalization and mortality [40], and the second included 22 geriatric randomized controlled trials, which showed that GA, combined with multidisciplinary interventions, improved survival, cognition and maintenance at home, and decreased the need for admission and institutionalization [41].

In older patients undergoing cancer treatment, three randomized trials have studied the impact of specific geriatric interventions on patients' outcomes [42-44]. In one randomized study conducted in 375 older post-surgical cancer patients, a specialized home care intervention was compared to usual care in terms of survival [42]. The intervention consisted of home visits and telephone calls by advanced practice nurses, who assessed and monitored patient physical, emotional and functional status over a 4-week period. The study demonstrated that the intervention increased survival when compared to usual post-operative follow up. Another randomized study has been conducted to evaluate the impact of nurse case management in improving medical care of 335 elderly women with breast cancer [43]. The intervention consisted of the services of a nurse case manager for 12 months after the diagnosis of breast cancer, who assessed functional status, cognition, mood and home safety to evaluate patient needs. Overall, a higher proportion of women in the intervention arm received breast-conserving surgery and radiation therapy, and underwent more breast reconstruction surgery. They were also more likely to recover normal functioning at 2 months after breast surgery. The third randomized study assessed the effect of geriatric inpatient and outpatient care on 99 older cancer patients compared to usual care [44]. The

intervention involved core teams that provided geriatric assessment and patient management. The study showed a positive effect of geriatric inpatient care on pain, emotional limitations and mental health. Moreover, the trial has evidenced no difference in hospital costs over 1 year between geriatric inpatient care and usual care.

Despite those promising findings, randomized controlled trials evaluating the efficacy of individualized interventions based on GA findings in older patients with cancer are still needed in order to measure important outcomes, namely: survival, hospitalizations, quality of life, treatment toxicities, completion of the planned cancer treatment, weight loss and functional decline.

### 1.3.2 Practical limitations of geriatric assessment in the oncology setting

Although it provides useful information, performing a complete geriatric assessment is time consuming, requiring up to 2 hours for the geriatrician to complete. It requires extensive involvement of geriatric teams, which are not always available. Further, the shortage of expert geriatricians in oncology can lead to delayed appointments for the assessment. The limited health care resources combined with the time and financial constraints makes difficult to integrate GA into the daily clinical practice [45]. On the other hand, an oncologist needs a brief assessment to specifically identify fit-appearing older individuals actually at risk for toxicities and poorer outcomes because of limited functional reserve and for whom a more complete geriatric assessment should be performed.

### 1.3.3 Frailty classifications based on GA parameters

Several attempts have been made in geriatric oncology to classify patients based on frailty components identified from GA and to propose these classifications as decision rules.

Such classifications are generally used as decision trees whose first step aims to determine patient health status, whether specific geriatric interventions are needed, and whether a complete GA is required. Secondly, decision-making based on health status assessment is proposed.

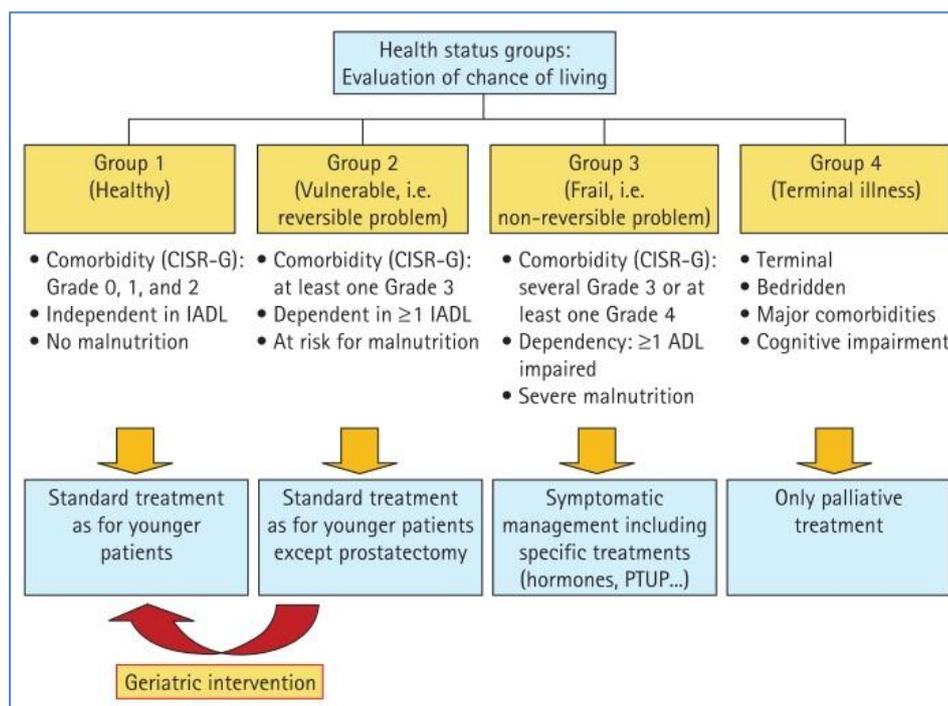
Balducci and Extermann proposed a definition for frailty for the older person with cancer (**Table 1.2**) [46], based on results from a study that identified functional impairment, comorbidity, and the presence of geriatric syndromes as predictors of mortality and institutionalization among male patients aged 65 years and older [47]. They complemented it by clinical expertise, including age 85 and older as a criteria for frailty, based on the observation that a majority of patients of that age presented functional and cognitive changes despite several studies previously reported that most patients above 85 should actually not be considered as frail [16]. Later, criteria were further modified to categorize patient health status into three groups, combining several components of GA, with recommendations for treatment plan [48]:

- **Group 1 (fit)**: patients who are functionally independent for ADL and without serious comorbidity, candidates for any form of standard cancer treatment, as long as their average life expectancy is longer than the life expectancy from cancer;
- **Group 2 (intermediate or vulnerable)**: patients who are functionally independent for ADL but may be dependent in one or more IADL and/or have 1 or 2 comorbidities and/or no geriatric syndromes. For this group decision could be either palliative or curative, but should be subject to some precautions, such as dose reduction of chemotherapy at first administration;
- **Group 3 (frail)**: patients aged  $\geq 85$  years, who are dependent for at least one ADL and/or have 3 or more comorbidities and/or at least 1 geriatric syndrome, mainly candidates for supportive care.

**Table 1.2.** Balducci criteria for frailty

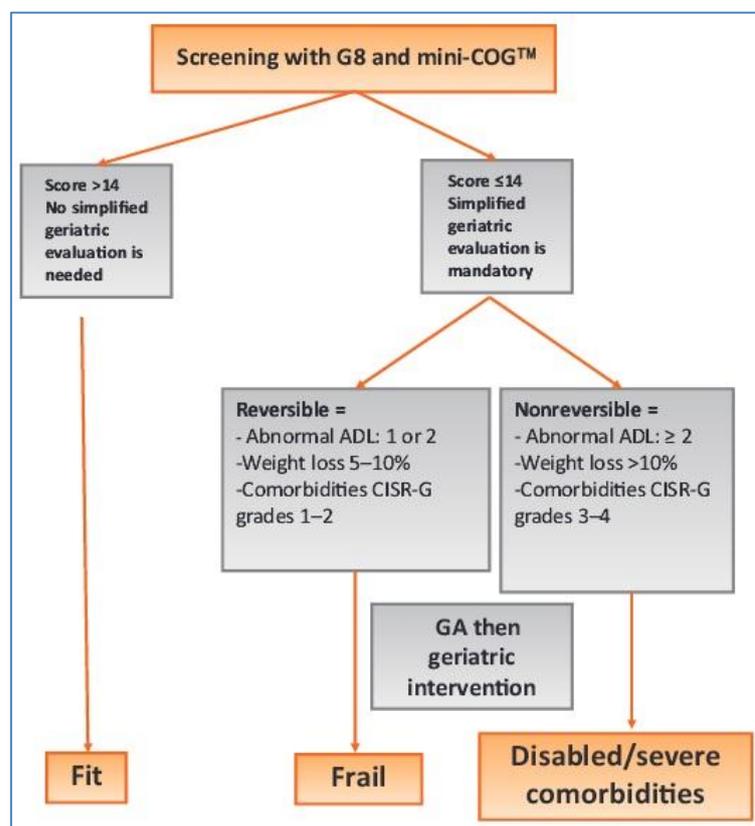
<b>The presence of any of these criteria establishes frailty</b>	
Age	>85
ADL (Activities of daily living)	Dependence in one or more
Comorbidity	One or more
Geriatric syndromes	One or more among: delirium, dementia, depression, osteoporosis, incontinence, $\geq 3$ falls/month, neglect and abuse, failure to thrive.

The SIOG task force has defined other classifications of patient health status, according to the groups defined by Balducci and Extermann with some adaptations for patients with prostate cancer. These guidelines classify patients into four groups (fit, vulnerable, frail and too sick) along with the implications for treatment, based on a systematic evaluation of comorbidities severity, dependence status, and nutritional status (**Figure 1.2**) [49].

**Figure 1.2.** Decision tree for treating patients with prostate cancer

CIRS-G, Cumulative Illness Rating Score-Geriatrics; ADL, Activities of Daily Living; IADL, Instrumental ADL; at risk for malnutrition: weight loss of 5-10% within the last 3 months; severe malnutrition: weight loss of >10% within the last 3 months.

Two other SIOG updated recommendations exist for the management of older patients with prostate cancer [50,51]. In the revised guidelines, evaluation of health status involves a stepwise process starting with screening using the Geriatric 8 (G8: screening tool to identify patients in need of GA, described later in this chapter) and the mini-COG™ (screening tool for cognitive impairment). This is followed, where indicated, by a simplified geriatric assessment and then, again when indicated, by full geriatric assessment, particularly when complex geriatric interventions are needed (**Figure 1.3**).



**Figure 1.3.** Decision tree to determine patient health status

Mini-COG™, Mini-COG™ cognitive test; ADL, Activities of Daily Living; CIRS-G, Cumulative Illness Rating Score-Geriatrics; GA, Geriatric Assessment

Regardless of the classification used to determine health status, the purpose is essentially to identify patients for whom individualized geriatric interventions may allow transitions between clinical states (i.e., from vulnerable to fit).

## 1.4 Screening tools for multidimensional health problems

Some authors have proposed a two-step approach for older patients with cancer: 1) health status screening; and 2) complete health status assessment for those patients identified in need of GA by the screening tool [52]. There are however important considerations to take into account to incorporate a screening tool into routine clinical practice (described below).

### 1.4.1 What does the term “screening tool” stand for?

According to the World Health Organization, *“screening is the presumptive identification of unrecognized disease or defects by means of tests, examinations, or other procedures that can be applied rapidly”* [53]. *“A screening test aims to be sure that as few as possible with the disease get through undetected (high sensitivity) and as few as possible without the disease are subject to further diagnostic tests (high specificity)”* [54]. A screening tool thus distinguishes apparently healthy people from those who probably have the disease or who may be at increased risk of a disease or condition. It is not intended to be diagnostic. Persons with positive or suspicious findings must be referred for further investigations, definite diagnosis and necessary treatment.

The factors that should be considered when assessing the efficacy and utility of a screening tool are summarized in **Table 1.3** [53,54].

**Table 1.3.** Factors to be considered for any screening tool.

<b>Factors</b>	
High sensitivity	Ability to classify as positive those people with the target condition.
High specificity	Ability to classify as negative those people without the target condition.
High positive predictive value	Extent to which subjects have the disease in those that give a positive test result.
High negative predictive value	Extent to which subjects are free of the disease in those that give a negative test result.
Acceptability	Extent to which those for whom the test is designed agree to be tested.  It should cause minimal disturbance to the subject in its performance.
Suitable for clinical practice	It should be simple, fast and easy to use.  Its implementation implies that it has been previously validated.
Low cost	The test should be as cheap as possible.

#### 1.4.2 Screening tools in geriatric oncology

As previously mentioned, notwithstanding its value, a complete geriatric assessment requires considerable time and human resources and is actually not required in all patients [55]. Thus, several screening tools have been developed to discriminate between fit older patients who are likely to tolerate standard therapy and vulnerable older patients who would benefit from complete GA and potential specific treatment tailoring based on its findings [56,57]. The ideal screening tool is easy to perform, requires little time, covers all the domains routinely assessed by geriatricians, and effectively separates fit from vulnerable patients [56]. A screening tool should recognize those patients who are able to benefit most for optimal cancer care plan, i.e. as much as younger adults, and those for whom the therapeutic risks overcome the potential benefits.

According to a recent review of Decoster et al. [56], 22 studies of older cancer populations compared a total of 14 screening tools with GA reporting on sensitivity and specificity. **Table 1.4** summarizes the geriatric domains and specific items assessed by these 14 screening tools. Not all domains are included evenly across the screening tools, as also showed by Hamaker et al. [57] in a previous systematic review of screening tools in older patients with cancer. While all of the screening tools assess functional status, only half of them assesses cognition and nutrition; six address comorbidity, age and/or hospitalizations, polypharmacy as well as social status, and the other domains are only evaluated by 5 (depression), 4 (health status) or 3 tools (fatigue). Overall, the average number of domains considered by a screening tool is 4, ranging from 1: Karnofsky Performance Status (KPS) [58] and Eastern Cooperative Oncology Group-Performance Status (ECOG-PS) [59], to 6: Geriatric-8 (G8) [60,61], Groningen Frailty Indicator (GFI) [62,63], Onco-geriatric screen (OGS) [64] and Gerhematolim [65,66]; these four tools evaluate functional status, cognition, nutrition and polypharmacy.

Regarding the individual items considered in each screening tool, mobility and weight loss were the most frequently used (by 7 screening tools), followed by prescription of daily medications (6 tools), although 2 different thresholds are considered (>3 drugs, 4 tools; >4 drugs, 2 tools). ADL impairments, namely, bathing (5 tools), shopping (5 tools) and dressing (4 tools) are also used as well as the fact of feeling sad or depressed (5 tools) and falls (4 tools). Other items (32) were found in 2 or 3 tools and the rest (24 items) were specific for a particular tool.

Relating to the high variability in the items used across screening instruments and their overall limited diagnostic performance (detailed below), it should be noticed that most tools, including the G8, were developed based on expert opinion and/or existing assessment tools that were originally validated to detect outcomes other than GA impairment.

**Table 1.4.** Domains and items evaluated by screening tools to identify older patients with cancer in need of GA

<b>DOMAINS</b>	<b>Items</b>	<b>Tool</b>	<b>Specifications according to each screening tool</b>		
<b>Autonomy / Functional status</b>	<i>ADL impairments</i>	Bathing	VES-13 aCGA OGS SAOP2 Gerhematolim	Yes (need help or not); No; Don't know. Receives or not assistance, or little. Does the patient need help? Yes; No. Does the patient need help? Yes; Occasionally; No. Without help: Yes; No.	
		Dressing	GFI OGS SAOP2 Gerhematolim	Be able to do without help? Yes; No. Does the patient need help? Yes; No. Yes; Yes but with help; No. Without help: Yes; No.	
	Toileting	GFI Gerhematolim	Be able to do without help? Yes; No. Without help: Yes; No.		
	Transferring	aCGA SAOP2	Moves into/out of bed/chair with, without assistance or does not get out of bed. Does the patient need help to get out of bed/chair? Yes; Occasionally; No.		
	Continence	aCGA SAOP2 Gerhematolim	Controls completely; occasional accidents; incontinent. Incontinent; occasionally; no incontinent. Problems of incontinence: Yes; No.		
	Feeding	Barber SAOP2 Gerhematolim	Unable to have a hot meal (any days): Yes; No. Yes; Yes but with help; No. Without help: Yes; No.		
	<i>IADL impairments</i>	Use the telephone	SAOP2 Gerhematolim	Yes; Yes but with help; No. Without help: Yes; No.	
		Shopping	GFI VES-13 aCGA OGS SAOP2	Be able to do without help? Yes; No. Difficulties: Yes (need help or not); No; Don't know. Without help; with help; unable. Does the patient need help? Yes; No. Yes; Yes but with help; No.	
		Preparing food	aCGA SAOP2	Without help; with help; unable. Yes; Yes but with help; No.	
		Housekeeping	VES-13 VES-13 aCGA	Difficulties for light housework: Yes (need help or not); No; Don't know. Difficulties for heavy housework: Not difficult; a little; some; a lot; unable to do. Without help; with help; unable.	
		Doing laundry	aCGA	Without help; with help; unable.	
		Mode of transportation	SAOP2 Gerhematolim	Drive a car: Yes; Have never driven; No. Public transport or car, without help: Yes; No.	
		Handling medications	OGS Gerhematolim SAOP2	Does the patient need help? Yes; No. Without help: Yes; No. Yes; Yes but with help; No.	
		Handling finances	VES-13 SAOP2 Gerhematolim	Yes (need help or not); No; Don't know. Yes; Yes but with help; No. Without help: Yes; No.	
		<i>Performance status</i>	Activity	KPS ECOG-PS	Normal activity; minor signs of disease; normal activity with effort; care for self; occasional assistance; considerable assistance; disabled; severely disabled; very sick, hospitalization; moribund. Fully active; restricted activity but ambulatory; up >50% of waking hours; confined to bed >50% of the day; completely disabled.

<i>Mobility/Falls/risk of falls</i>	Mobility	G8	Bed/chair bound; gets out of bed/chair but does not go out; goes out.
		GFI	Walking around/outside the house or to the neighbors without assistance: Yes; No.
		VES-13	Walking a quarter of a mile (Not difficult; a little; some; a lot; unable to do);
		VES-13	Walking across the room (Difficulties: Yes (need help or not); No; Don't know)
	Falls	Barber	Confined to home through ill health: Yes; No.
		fTRST	Reduced mobility: Yes; No.
		SAOP2	Use cane, walker or wheelchair: Yes; Yes occasionally; No.
		OGS	In the past 3 months: Yes; No.
		fTRST	In the past 6 months: Yes; No.
		SAOP2	In the past year: Yes; No.
<i>Physical function limitations</i>	Physical fitness	Gerhematolim	In the past 3 months: Yes; No.
		GFI	Scale of 0 to 10 (From very bad to very good)
	Reduced gait speed	Fried	Low activity: <20% kcal/week
		Fried	<20% walking time/15 feet (by gender, height)
	Timed up and go	Gerhematolim	<20 seconds.
		VES-13	Not difficult; a little; some; a lot; unable to do.
	Stooping, crouching or kneeling	VES-13	Not difficult; a little; some; a lot; unable to do.
		VES-13	Not difficult; a little; some; a lot; unable to do.
	Reaching or extending arms above shoulder level	VES-13	Not difficult; a little; some; a lot; unable to do.
		VES-13	Not difficult; a little; some; a lot; unable to do.
Inability to rise from a chair five times without using the arms	SOF	Yes; No.	
	VES-13	Not difficult; a little; some; a lot; unable to do.	
Writing, or handling and grasping small objects	VES-13	Not difficult; a little; some; a lot; unable to do.	
	VES-13	Not difficult; a little; some; a lot; unable to do.	
<i>Neurosensory deficits</i>	Poor vision	GFI	Experience problems in daily life due to: Yes; No.
		Barber	Yes; No.
		ISAR	Yes; No.
	Poor hearing	GFI	Experience problems in daily life due to: Yes; No.
		Barber	Yes; No.
		Barber	Yes; No.
<b>General health status</b>	Self-rated health status (compared to other people of same age)	G8	Not as good; does not know; as good; better.
		VES-13	Poor; fair; good; very good; excellent.
	Is there anything about your health causing you concern or difficulty?	SAOP2	Scale of 1 to 10 (10 is excellent, 1 is poor)
		Barber	Yes; No.
Present quality of life	SAOP2	Scale of 0 to 10 (From worst to best life)	
	SAOP2	Yes; No.	
<b>Fatigue / Weakness</b>	Poor energy / fatigue	SOF	Yes; No.
		Fried	Self-reported fatigue.
	Hand grip strength	Fried	Low handgrip strength: <20% (by gender, body mass index)
Lifting/carrying heavy objects	VES-13	(10 pounds): Not difficult; a little; some; a lot; unable to do.	
	VES-13	(10 pounds): Not difficult; a little; some; a lot; unable to do.	
<b>Cognition</b>	Neuropsychological problems (dementia or depression)	G8	Severe dementia or depression; Moderate dementia or depression; no problems.
		GFI	Diagnosed with dementia: Yes; No
		Gerhematolim	Learn 3 names: lemon, key, ball.
	Registration	ISAR	Yes; No.
		GFI	Memory complains: Yes; No.
	Memory problems	OGS	Memory loss: Yes; No (several episodes per day)
		fTRST	Yes; No.
		OGS	Yes; No (not remember the date)
	Poor recall or not oriented	Gerhematolim	Recall 3 names learned earlier.
		aCGA	Exercise (subtract 7 from 100 5 times): 5 points
	Attention and calculation	Gerhematolim	Exercise (subtract 7 from 100 5 times): 5 points
		aCGA	Exercise: 1 point
	Reading	aCGA	Exercise: 1 point
aCGA		Exercise: 1 point	
Writing	aCGA	Exercise: 1 point	
	aCGA	Exercise: 1 point	
Copying	aCGA	Exercise: 1 point	
	aCGA	Exercise: 1 point	

<b>Depression</b>	Experience emptiness	GFI	Yes; No
		aCGA	Yes; No.
	Feel helpless	Gerhematolim	Yes; No.
		aCGA	Yes; No.
	Feel worthless	Gerhematolim	Yes; No.
		aCGA	Yes; No.
	Feel happy/sad or depressed	GFI	Recently felt downhearted or sad: Yes; No.
		aCGA	Yes; No.
		OGS	Yes; No.
		SAOP2	Yes; No.
Gerhematolim		Yes; No.	
Loss of interest in usual activities	OGS	Yes; No.	
	SAOP2	Yes; No.	
Miss people around	GFI	Yes; No.	
Feeling nervous or anxious	GFI	Recently : Yes; No	
<b>Nutrition</b>	Decrease in food intake, appetite loss	G8	Severe anorexia; moderate anorexia; no anorexia.
		SAOP2	In the last 3 months: Yes; No.
	Change in type of foods able to eat	SAOP2	Yes; No.
		G8	In the last 3 months: >3 kg; does not know; 1-3 kg; no WL.
	Weight loss (WL)	GFI	6 kg in the past 6 months or 3 kg in 1 month: Yes; No.
		SOF	≥ 5% the preceding year (intentionally or unintentionally)
		Fried	≥ 10 pounds during the preceding year: Yes; No.
		OGS	≥ 10% in the past 6 months: Yes; No.
		SAOP2	≥ 5 pounds in the past 6 months: Yes; No.
		Gerhematolim	> 5% in the past 6 months: Yes; No.
Body Mass Index – BMI (kg/m <sup>2</sup> )	G8	<18.5; 18.5≥ BMI <21; 21≥ BMI<23; ≥23.	
	Gerhematolim	< 21; ≥21.	
Albumin (g/l)	OGS	≤35; >35.	
	Gerhematolim	≤30; >30.	
<b>Comorbidity/Age/ Hospitalizations</b>	Comorbid conditions	OGS	Unstable or untreated?: Yes; No.
	Age	G8	<80 years; 80-85 years; >85 years
		VES-13	75-84 years; ≥85 years
	Hospitalizations	fTRST	In the last 3 months: Yes; No.
		Barber	During the past year: Yes; No.
<b>Polypharmacy</b>	Prescription drugs per day	ISAR	In the last 6 months
		G8	> 3
		GFI	≥ 4
		fTRST	≥ 5
		ISAR	> 3
		OGS	> 4
<b>Social status</b>	Lives alone	Gerhematolim	> 3
		fTRST	Yes; No.
		Barber	Yes; No.
	Available caregiver/relative	Gerhematolim	Yes; No.
		Barber	You could call on for help: Yes; No.
		fTRST	Yes; No.
	Assistance at home	Gerhematolim	Relative of professional: Yes; No.
		SAOP2	Someone who could take care of you if necessary: Yes; No.
		ISAR	Before and after emergencies admission: Yes; No.
		Barber	Do you depend on someone for regular help: Yes; No.
Self-manage in an emergency	Gerhematolim	Yes; No.	
Feel left alone	GFI	Yes; No.	
Able to pay for medications	SAOP2	Yes; No.	

ADL, Activities of Daily Living; IADL, Instrumental ADL; VES-13, Vulnerable Elders Survey; aCGA, Abbreviated Comprehensive Geriatric Assessment; OGS, Onco-geriatric screen; SAOP2, Senior Adult Oncology Program 2 screening; GFI, Groningen Frailty Indicator; KPS, Karnofsky Performance Status; ECOG-PS, Eastern Cooperative Oncology Group-Performance Status; fTRST, Flemish version of the Triage Risk Screening Tool; ISAR, Identification of Seniors At Risk; SOF, Study of Osteoporotic Fractures Index.

## Diagnostic performance of screening tools

Hamaker et al. [57] were the first to systematically review the diagnostic performance of several geriatric screening tools for detection of impairments on a complete geriatric assessment in older patients with cancer. It included 14 studies evaluating 7 different tools and found limited discriminative power with a wide range in sensitivity and specificity, from 25% (Fried criteria) to 92% (G8 and fTRST[cutoff of  $\geq 1$ ]), and from 39% (G8) to 100% (fTRST[cutoff of  $\geq 2$ ]), respectively. A task force convened by the SIOG conducted a more recent systematic review from which 22 studies comparing 14 screening tools with GA were identified, and reported similar results [56]. Since then, 9 other studies found in the literature [67-75] have evaluated 7 screening tools in older patients with cancer (among those identified by the two systematic reviews of Hamaker et al. and Decoster et al.). Annex 1 (**Table A1.1**) details the characteristics (number of patients, study population and number of GA items) and the results (sensitivity and specificity according to the cutoff for abnormal GA) of these studies. The number of domains (or items) considered to define an abnormal GA ranged from 4 to 10. Ten studies used a cutoff for GA impairment of  $\geq 1$ , 19 used a cutoff of  $\geq 2$  and 3 studies reported performance values according to both thresholds. This large variation of the definition of an abnormal GA makes inter-study comparison difficult.

The most studied screening tools by order of decreasing number of patients are the G8 (n=4630), the VES-13 (n=3303), the fTRST (n=1212) and the GFI (n=666). **Table 1.5** summarizes the diagnostic properties in terms of sensitivity and specificity of the 14 screening tools identified across the 32 studies. In brief, although many tools have been described, none combines both appropriate sensitivity and specificity for predicting an abnormal GA.

**Table 1.5.** Diagnostic performance of screening tools to identify older patients with cancer in need of geriatric assessment

Screening tool	No. of items	C-off <sup>a</sup>	No. of Studies	GA $\geq 1$ impairment		No. of Studies	GA $\geq 2$ impairments	
				Se (%)	Sp (%)		Se (%)	Sp (%)
KPS (scale: 0-100) [58]	11	$\leq 80$	1	29	44	1	78	91
ECOG-PS (scale: 0-5) [59]	5	$\geq 1$	-	-	-	1	94	55
		$\geq 2$	-	-	-	2	59 (64-64)	90 (81-99)
Barber tool [76]	9	$\geq 1$	-	-	-	2	67 (59-74)	59 (39-79)
fTRST [77]	5	$\geq 1$	-	-	-	3	91 (59-92)	50 (42-86)
		$\geq 2$	-	-	-	2	66 (64-67)	90 (80-100)
ISAR [78]	6	$\geq 2$	1	70	10	-	-	-
VES-13 [79]	13	$\geq 3$	5	68 (60-87)	71 (62-81)	13	60 (15-88)	81 (64-100)
Fried criteria [22]	5	$\geq 3$	2	28 (25-31)	97 (96-98)	2	45 (37-52)	89 (86-92)
		$\geq 1$	1	87 (81-92)	49 (38-60)	-	-	-
GFI [62,63]	15	$\geq 4$	-	-	-	5	64 (39-79)	86 (71-87)
aCGA [80,81]	4 <sup>b</sup>	$\geq 1$	-	-	-	3	79 (51-84)	86 (59-97)
OGS [64]	5 <sup>b</sup>	$\geq 1$	1	88 (80-93)	44 (28-63)	-	-	-
SAOP2 [82,83]	23	$\geq 2$	-	100	40	-	-	-
SOF [84]	3	$\geq 2$	1	-	-	1	89 (85-93)	81 (73-88)
Gerhematolim [65,66]	7 <sup>b</sup>	$\geq 2^c$	-	-	-	1	95	87
G8 [60,61]	8	$\leq 14$	6	86 (65-90)	62 (3-100)	12	87 (38-97)	55 (29-79)

Data are presented with median and range when the screening tool is evaluated in more than one study or with value and 95% confidence interval, otherwise (if reported).

Abbreviations: C-off, Cut-off; GA, geriatric assessment; Se, sensitivity; Sp, specificity; KPS, Karnofsky Performance Status; ECOG-PS, Eastern Cooperative Oncology Group-Performance Status; fTRST, Flemish version of the Triage Risk Screening Tool; ISAR, Identification of Seniors At Risk; VES-13, Vulnerable Elders Survey; GFI, Groningen Frailty Indicator; aCGA, Abbreviated Comprehensive Geriatric Assessment; OGS, Onco-geriatric screen; SAOP2, Senior Adult Oncology Program 2 screening; SOF, Study of Osteoporotic Fractures Index; G8, Geriatric-8.

<sup>a</sup> Cutoff for the screening tool to be abnormal.

<sup>b</sup> Number of domains to be evaluated by different items

<sup>c</sup> or functional impairment

The G8 (**Table 1.6**), a screening tool specifically developed for older patients with cancer [61,85], is among the most sensitive tools available for detecting patients with impaired GA findings but lacks specificity in its original version. With the aim of developing a more accurate screening tool, Pottel et al. [86] performed exploratory analyses combining the VES-13 and the G8 screening tools. Even if the combined score improved accuracy in this study of older patients with head and neck cancers, this was not confirmed in a later study [74]. Other direct modifications of the G8 have recently been proposed based on literature review and Delphi consensus [87]. Several domains

associated with frailty were considered for the modifications, with one alternative (IADL-modified G8) showing a slightly better specificity (67% vs. 64%,  $p < 0.05$ ) but very similar sensitivity (77% vs. 76%,  $p = 0.53$ ), thus still not meeting existing needs for geriatric practice.

**Table 1.6.** G8 screening tool

<i>Items</i>	<i>Score</i>
<b>1</b> Has food intake declined over the past 3 months due to loss of appetite, digestive problems, chewing or swallowing difficulties?	0: Severe decrease in food intake 1: Moderate decrease in food intake 2: No decrease in food intake
<b>2</b> Weight loss during the past 3 months	0: >3 kg 1: patient does not know 2: 1-3 kg 3: No weight loss
<b>3</b> Mobility	0: Bed or chair bound 1: Able to get out of bed/chair but does not go out 2: Goes out
<b>4</b> Neuropsychological problems	0: Severe dementia or depression 1: Mild dementia or depression 2: No neuropsychological problems
<b>5</b> Body mass index - BMI (weight in kg/height in m <sup>2</sup> )	0: <18.5 1: 18.5 ≥ BMI <21 2: 21 ≥ BMI <23 3: ≥23
<b>6</b> Takes more than 3 prescription drugs per day	0: Yes 1: No
<b>7</b> Compared to other people of the same age, how does the patient considerate his or her health status?	0: Not as good 0.5: Patient does not know 1: As good 2: Better
<b>8</b> Age	0: >85 years 1: 80-85 2: <80
<b>Total</b>	___ / 17

The total score is the sum of the scores on each of the 8 items.

A total score ≤14 is considered abnormal and warrants a full geriatric assessment in the two-step approach.

## Screening tools with prognostic value for other outcome measures

Several studies have demonstrated an association of screening tools with outcomes other than results from the GA. The G8 and the fTRST have shown to be predictive for functional decline [88]. Associations with chemotherapy-related toxicity have been shown for G8 [89], VES-13 [89,90], and GFI [91]. Finally, five screening tools have been associated with overall survival: G8 [88,92], VES-13 [93], fTRST [88], GFI [94,95], and Fried Frailty Criteria [96]. It is important to emphasize that even if screening tools provide important information about treatment-related toxicity, risk of functional decline, and overall survival, they should not replace GA.

### Summary

- The geriatric assessment has proved to be useful as an aid to medical decision making for older patients with cancer, but remains particularly **time-consuming** given the multiple components and scales required for its completion (nutritional status, autonomy, cognitive functions, comorbidities, etc.).
- Given that the majority of oncology patients are over age 65, in most clinical oncology settings it is not practical to perform a GA for all older patients.
- Therefore, several shorter scales have been developed and proposed as screening tools for a two-step approach. However:
  - 4) their diagnostic performance remains **insufficient** for clinical practice;
  - 5) most instruments have been developed exclusively on the basis of experts' opinions without any **specific statistical / psychometric development**; and
  - 6) a wide variability of criteria have been used to define "frailty" as the gold standard, without **investigating the influence of such variations** on the diagnostic properties and/or the concepts being actually measured.

## 1.5 Research objectives

### 1.5.1 Primary objective

The main objective of this thesis was to develop a screening tool that would enhance the early identification of older cancer patients with a geriatric risk profile requiring a complete GA, in order to ensure a more appropriate management of the cancer itself and related impairments. In particular, we addressed to optimize the G8 screening tool following a detailed step-by-step statistical analysis, by improving current items and adding potentially useful new ones while targeting high discriminative power, usability and clinical relevance.

Other steps included the internal (by bootstrap technics) and external (in one independent population) validation of its diagnostic performance, which is crucial to assess the stability of the model.

### 1.5.2 Secondary objectives

This thesis stated two secondary objectives:

1. to measure and compare the effect of varying gold standard definitions on the diagnostic performance of the G8 and the modified G8 screening tools, both specifically developed for older patients with cancer, and
2. to assess and compare the prognostic value for survival of both screening tools, the original G8 and its optimized version, the modified G8 in a large cohort of older patients with cancer, overall and by tumor site.

## CHAPTER 2

---

### **General methodology**

## 2.1 Study design

We analyzed data from two different surveys: 1) the Elderly CAncer PAlients study (ELCAPA), used to develop the screening tool modified G8, to assess its prognostic value and to examine the variability of its diagnostic performance under multiple gold standard definitions; and 2) the ONCODAGE study, used to externally validate the modified G8 in older patients with cancer. Each survey is described in detail below.

### 2.1.1 The ELCAPA Prospective Cohort Study

ELCAPA study is a prospective open-cohort survey with consecutive enrolment of patients aged 70 years and older with a newly diagnosed histologically confirmed solid or hematological cancer at all stage, who were referred by an oncologist, surgeon, radiotherapist, or other specialist to one of 17 geriatric oncology clinics in teaching hospitals in the Paris urban area, France. Study inclusion occurs on the day of the first geriatric-oncology visit in which a multidimensional GA is performed. By February 2017, 2952 patients had been included. **Figure 2.1** shows the centers that are currently recruiting patients.

For the first work of this thesis (development of the screening tool modified G8), only 2 centers were opened: Henri Mondor Hospital (Créteil) and Paul Brousse Hospital (Villejuif).

The work analyzing the prognostic value of the modified G8 screening tool and its temporal validation, additionally included patients from 4 other centers: Créteil Intercommunal Hospital Center, Bretonneau Hospital (Paris), Louis-Mourier Hospital (Colombes) and Curie Institut (Paris).

Finally, the study analyzing the effect of varying gold standard definitions on the diagnostic performances of the *original* and *modified G8* screening tools, included patients from 10 centers, the 6 already mentioned, plus René Huguenin Center (Saint-Cloud), Broca-Cochin Hospitals (Paris), Ambroise-Paré Hospital (Boulogne-Billancourt) and the European Hospital Georges-Pompidou (Paris).

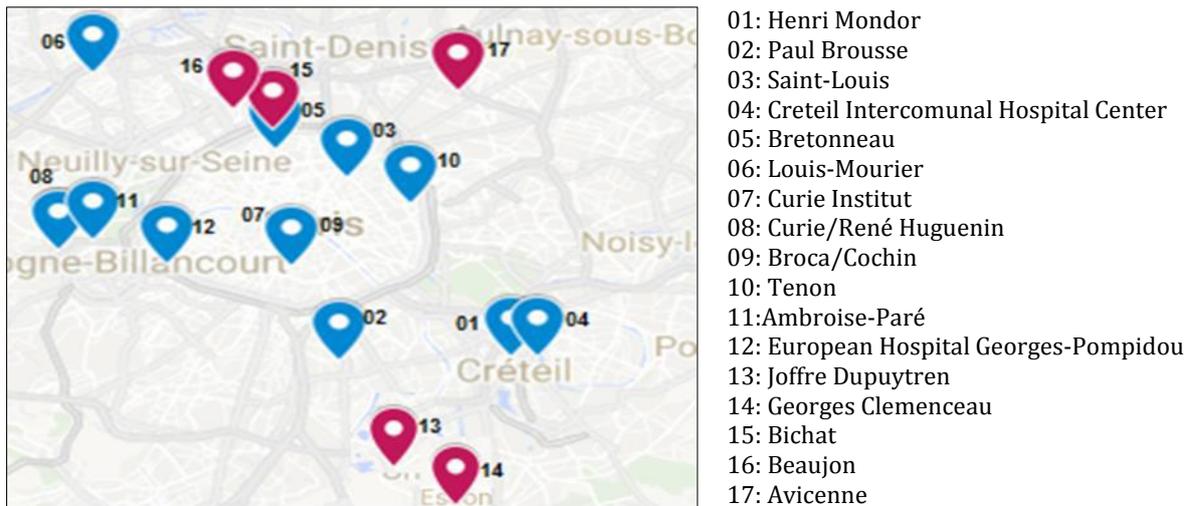


Figure 2.1. ELCAPA study centers (February 2017)

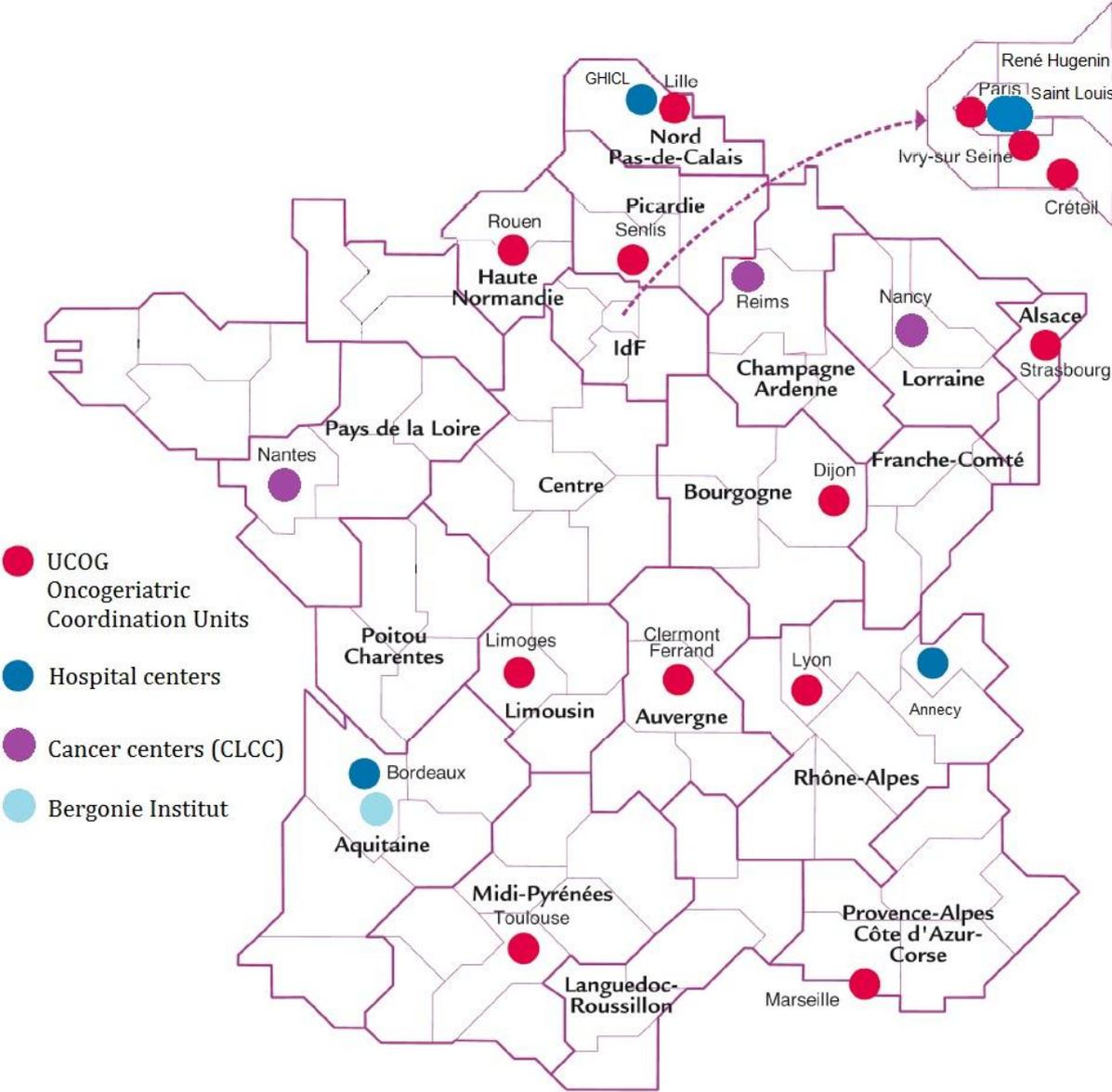
Informed consent was obtained from all study patients prior to inclusion. The protocol was approved by the appropriate ethics committee: “Comité de Protection des Personnes” [CPP] Ile-de-France I, Paris, France. (Clinical Trial registration: NCT02884375).

### Survey objectives

The general objectives of the survey are 1) to assess the role of GA for decision making process for older patients with cancer, 2) to identify geriatric and oncologic factors associated with overall survival, treatment feasibility, toxicities and morbidities, 3) to develop and/or validate screening tests for frailty in geriatric oncology (purpose for which my thesis was conducted) and 4) to develop and validate frailty classifications.

### 2.1.2 The ONCODAGE Prospective Multicenter Cohort Study

This national multicenter study was designed in response to a French National Cancer Institute (INCa) call for proposal to identify a geriatric screening tool for detecting risk in cancerology. Patients were recruited from 23 health care facilities including 15 INCa accredited Regional Coordination Units for Geriatric Oncology. **Figure 2.2** presents the centers involved in the ONCODAGE project.



**Figure 2.2.** ONCODAGE study centers (March 2010)

GHICL, Groupement des Hôpitaux de l'Institut Catholique de Lille; CLCC, Centres de Lutte Contre le Cancer.

Eligible patients were 70 years and older with histologically confirmed cancer from various tumor sites. Patients were included either before any first-line treatment, or between any two steps of a pre-defined first-line treatment sequence (chemotherapy, endocrine therapy, targeted treatment, surgery or radiotherapy). Patients with known central nervous system metastases were excluded.

All patients provided written informed consent prior to enrollment. The protocol was approved by institutional review boards and ethics committees, and was conducted in accordance with the Declaration of Helsinki, Good Clinical Practices (Clinical Trial registration: NCT00963911).

### *Survey objectives*

The primary objective of this study was to validate the screening tool G8 to identify older cancer patients requiring geriatric assessment. The secondary objectives are 2) to validate the French version of the screening tool "Vulnerable Elders Survey (VES-13)", 3) to evaluate the merits of the screening tools (G8 and VES-13), 4) to assess the screening tool in specific populations, 5) to assess the number and type of interventions proposed after thorough geriatric assessment, and 6) to compare the two new tools (G8 vs. VES-13).

## 2.2 Participants

According to each objective based on ELCAPA study, we used data from patients enrolled in the survey at the time of the analysis and for whom required information for the specific analysis was available in the electronic database, as follows:

1. For the ***development study*** (ELCAPA-07), we included patients for whom complete G8 and GA information was available. For the ***temporal validation study***, all patients with complete GA were included. Patients enrolled from January 2007 to October 2012 were considered as the training set, whereas patients subsequently recruited until July 2014 were considered as the validation set.
2. For the ***prognosis value study***, we retained data for patients recruited between January 2007 and April 2014 with follow-up data and screening scores information available.
3. For the ***gold standard definitions study***, we included patients enrolled between January 2007 and June 2015 with complete data on each of the six reference standard definitions tested and available G8 and modified G8 data.

For the ***external validation study*** we included eligible and evaluable patients based on the original ONCODAGE protocol, recruited from August 2008 to March 2010, for whom complete GA information and modified G8 data was available.

**Figure 2.3** summarizes the flow diagram of patients for each study

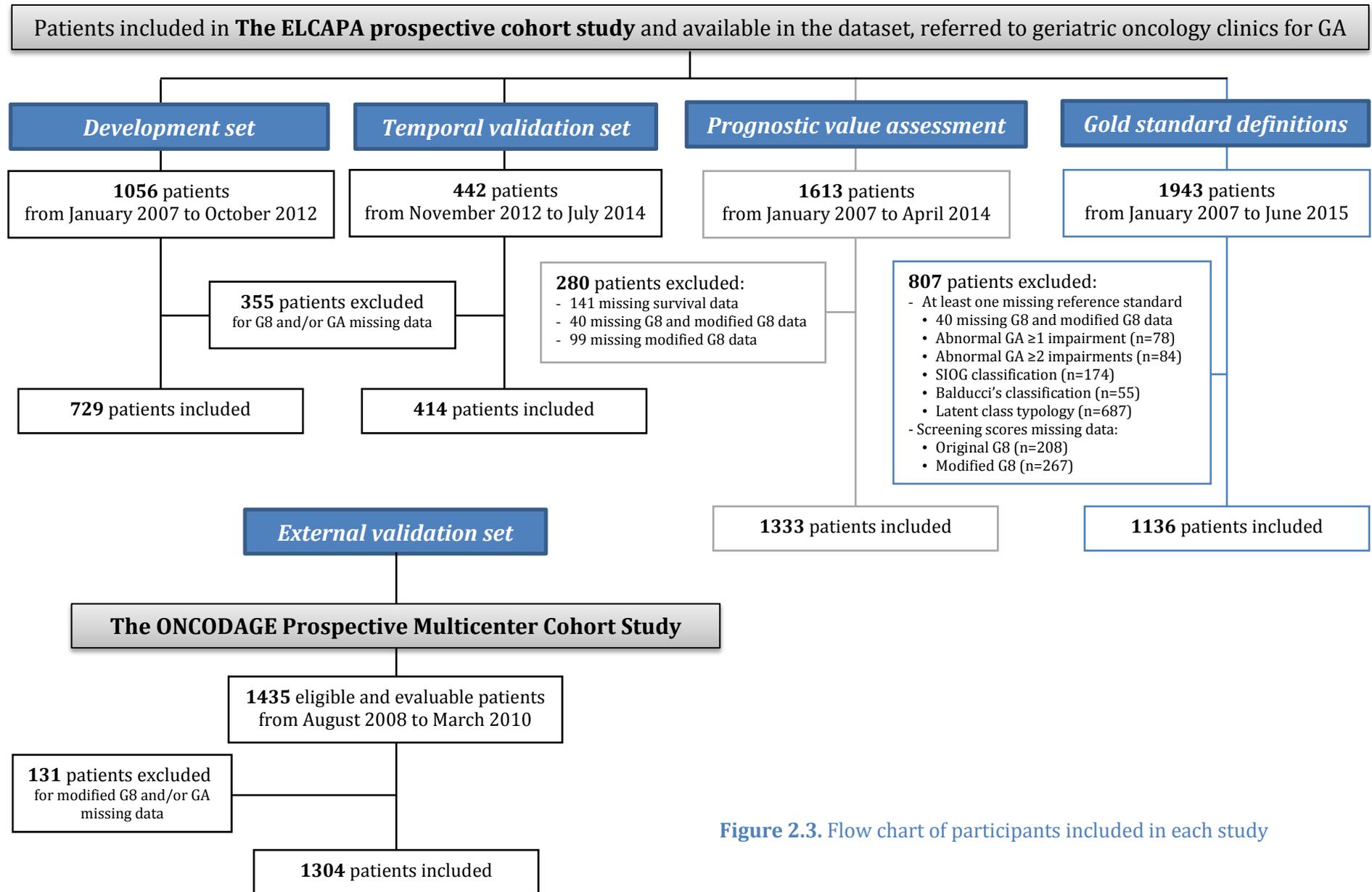


Figure 2.3. Flow chart of participants included in each study

## 2.3 Data collection

For the two cohorts analyzed in this manuscript, similar procedure of data collection was performed, with only some variations that will be mentioned in the appropriate section.

During the oncologic-geriatric visit of approximately 120 minutes, a complete clinical examination and a multidimensional GA was performed. In the ELCAPA study, a senior geriatrician specialized in oncology completed every test and collected all patient information prospectively on a standardized case report form. In the ONCODAGE study, 6 of the 7 instruments of the GA was completed by a nurse (ADL, IADL, MNA, TUG, MMSE and the GDS-15), and the geriatrician rated comorbidity on the CIRS-G. **Annex 2** presents the tests used for the GA.

In both cohorts, at the end of the GA the geriatrician proposed necessary interventions for overall patient management. Then, a multidisciplinary meeting was held to determine the best treatment strategy.

The following information was recorded in each cohort:

- ***Socio-demographic and cancer-related characteristics:*** age, gender, marital status, place of residence, the fact of living alone at home, having support at home and/or relatives available, tumor site, metastatic disease status (M0: absence of distant metastases; M1: presence of distant metastases; Mx: metastatic status not assessable) and anticipated therapeutic strategy.

– **Results from the multidimensional GA:** the main domains evaluated during the geriatric consultation were: 1) *functional status*, assessed by the ADL [97], IADL [98] and ECOG-PS [59]; 2) *mobility*, assessed by the Timed Up-and-Go test [99] in the 2 cohorts, plus history of fall(s) in the past 6 months and single-leg stance time (risk of falls) in ELCAPA; 3) *nutritional status*, assessed by the MNA [100] in the 2 cohorts, plus other parameters such as weight loss, BMI and low albumin in ELCAPA ; 4) *cognitive status*, assessed by the MMSE [101] in the 2 cohorts, plus history of cognitive disorders in ELCAPA; 5) *mood*, assessed by the mini-GDS [102] and the DSM-IV criteria (Diagnostic and Statistical Manual of Mental Disorders) [103] in ELCAPA and by the GDS-15 [104,105] in ONCODAGE; 6) *comorbidities*, assessed by the CIRS-G [106,107]. Additional information was collected in both cohorts: the number of daily prescribed medications and biological data such as albumin, hemoglobin, creatinine, C Reactive Protein and Low-Density Lipoproteins. Other comorbidities and geriatric syndromes were additionally reported in ELCAPA, namely diabetes, hypertension, heart failure, coronary heart disease, complete arrhythmia with atrial fibrillation, chronic renal failure, chronic respiratory failure, neurological deficit, urinary and/or fecal incontinence, vision and hearing impairments and asthenia.

– **Screening tool G8:** in ELCAPA, the instrument was scored using the available information in the database and with published guidelines for scoring [61]. In ONCODAGE, at the first visit after enrolment, patients completed the G8 test with a nurse, a clinical research assistant or a physician. As described in **Table 1.6** (Chapter 1), the total score for the G8 items (nutritional data, weight loss, motor skills, neuropsychological status, body mass index, medication, self-rated health status, and age) ranges from 0 to 17, a higher score indicating better health status. Following Bellera et al. [61], an abnormal G8 score was defined as  $\leq 14$ .

- **Interventions proposed by the geriatrician after GA:** modification or adaptation of the initial anticancer treatment and/or other therapeutic modifications, nutritional support, social support, nursing, kinesitherapy, neuropsychological support, further investigations such as imaging and biological tests, and referrals to other specialists in other fields (i.e. cardiologist, pneumologist, gastroenterologist, endocrinologist, ophthalmologist.).
- **Vital status follow-up data:** vital status was identified in medical charts or at the public records office. In ELCAPA, information on vital status was collected at 6 months, then every year through 5 years of follow-up. In ONCODAGE, vital status was collected at 2 time points: at 1-year and 5-year follow-up.

## 2.4 Data Analysis

For each study, patient characteristics are described with number (percentage) for qualitative variables and median (interquartile range, IQR) for quantitative variables. Comparisons of baseline characteristics of included and non-included populations were performed by the t-test or Wilcoxon Mann–Whitney test for quantitative variables and the chi-square or Fischer’s exact test for qualitative variables, as appropriate. Specific statistical analyses for each study are described in the corresponding section.

Most analyses were performed using Stata v13 (StataCorp, USA) at a two-tailed  $P < 0.05$  level. Exceptions included the Firth’s procedure in survival analysis and the calibration plots in the external validation, which were performed using R v3.3.0 (R Foundation, Austria; packages ‘coxph’ and ‘rms’).

## CHAPTER 3

---

### **Development of the screening tool modified G8**

### **3.1 Introduction**

Elderly patients with cancer are a very heterogeneous population because of differences in their health status that may vary considerable and that leads to different treatment responses and toxicity effects. Therefore, assessing each individual patient is crucial to provide optimal care. As previously mentioned, a multidimensional GA is recommended in this population, as it produces an inventory of health problems and an evaluation of physical, psychosocial, and functional capabilities [26]. The GA has proved useful for characterizing health and functional impairments potentially associated with oncological outcomes [28,35,55,108]. Nevertheless, it requires considerable time and human resources and not required in all patients [55].

Screening tools have therefore been developed to discriminate between fit older patients and vulnerable patients who would benefit from specific treatment tailoring based on findings from a complete GA [56], but none combines appropriate sensitivity and specificity for predicting an abnormal GA [56,57]. The G8 screening tool [61,85] has one of the highest sensitivities but it lacks specificity.

Here, our objectives were to evaluate the performance of the G8 in identifying older cancer patients likely to have abnormal GA and to determine whether modifications to this instrument might improve this performance. We used a systematic statistical approach to simultaneously test the original G8 items and a set of additional, potentially relevant items in a large cohort of older patients with cancer.

## 3.2 Methods

### 3.2.1 Study population

The ELCAPA prospective cohort is described in Chapter 2 (General methodology).

**Figure 2.3** presents the flow chart of participants included in this study.

### 3.2.2 GA Reference Procedure

The GA consisted of a set of seven validated tests covering a variety of important health domains in older cancer patients [109] and consistent with the questionnaires and thresholds used in the primary G8 validation study [61]. Abnormal GA was thus defined as an impaired score on at least one of the following tests: ADL  $\leq 5/6$ , IADL  $\leq 7/8$ , MMSE  $\leq 23/30$ , Mini GDS  $\geq 1$ , MNA  $\leq 23.5/30$ , CIRS-G; at least one comorbidity grade 3 or 4, and TUG  $>20$  seconds. Sensitivity analyses were performed to test the robustness of the results, by using an alternative cutoff of  $\geq 2$  impaired tests to define abnormal GA or by omitting the CIRS-G or ADL/IADL from the reference GA or by adding a social domain, as recommended by the SIOG [28], using the two cutoff values ( $\geq 1$  or  $\geq 2$  impaired tests).

### 3.2.3 Candidate Items

We identified 22 candidate items for a modified G8, based on both the literature and clinical expertise, to avoid the overfitting to the training data set seen when item selection relies solely on statistical significance [110]. In addition to the G8 items (depression/dementia, body mass index [BMI], anorexia, weight loss, age, medications [ $>3$  per day], mobility, and self-rated health status), we selected 14 items routinely collected during geriatric evaluations and known to be clinically relevant for assessing older cancer patients: asthenia, incontinence, fall risk (single-leg stance time  $<5$  seconds), history of fall(s) in the past 6 months, Eastern Cooperative Oncology Group

Performance Status (ECOG-PS), gender, living alone at home, metastatic status, and a selection of six comorbidities (diabetes, hypertension, heart failure and/or coronary heart disease [CHD], complete arrhythmia with atrial fibrillation [CAAF], chronic renal failure, and chronic respiratory failure).

### 3.2.4 Statistical Analysis

First, we performed a descriptive univariate analysis to assess associations between candidate items and the reference GA, using  $\chi^2$  or Fisher's exact tests as appropriate and estimating the crude odds ratios (ORs) with their 95% confidence intervals. A Firth's penalization procedure was applied to account for the small numbers among some categories or for perfectly prediction of an abnormal GA [111]. Receiver-operating curve (ROC) analysis was performed for the number of medications to assess alternative cutoffs to the one used in the original G8.

Second, we used MCA to investigate correlations between candidate items and identify redundancies across conceptually close qualitative variables, thus helping to decide which variables should be combined, dichotomized, or omitted [112]. Derived from the variables, MCA identifies common factors (or dimensions) that best represent all variables (the smallest number of dimensions that account for the largest proportion of the total variance or inertia) and can be interpreted as components of the health status. The major categories are those that have the highest quality of representation of factors (squared correlations between categories and the dimensions) and contribute most to their formation (contribution above the average:  $>1/\text{total number of categories}$ ). The graphical representation of variable categories according to their coordinates illustrates their contributions. Supplementary or illustrative elements can also be projected onto the factors for interpretation and comparison purposes.

In this regard, an abnormal GA was used as a supplementary variable without contribution to the formation of factors.

Remaining candidate items were then entered into a multivariate logistic regression model, using a stepwise backward procedure to sequentially remove items based on a  $p < 0.05$  level until the final model was obtained. Regression coefficients were considered for use as weights to compute the final score. We rescaled (multiplied) and rounded them to the closest integer, using the algorithm described by Cole to find the optimal solution that both improved simplicity of use in the clinical setting and preserved initial model accuracy [113]. This algorithm consists thus in finding the smallest common multiplier,  $k$ , which permits each estimated coefficient to be transformed into an integer without too much loss of precision. Model discrimination was assessed by the area under the ROC curve (AUROC) and calibration by the Hosmer-Lemeshow  $\chi^2$  test.

We internally validated our model using bootstrapping procedures with 300 replications to estimate the amount of optimism in our measurement of model discrimination and to compute the bias-corrected AUROC accordingly [114]. The modified G8 was finally applied to the temporal validation set population in which AUROC, sensitivity, specificity, positive predictive values (PPV), negative predictive values (NPV), positive likelihood ratio (LR+), and negative likelihood ratio (LR-) were calculated. While predictive values significantly dependent on the prevalence of the disease in the population studied, the likelihood ratios are independent of prevalence and can be used as alternatives to quantify diagnostic accuracy. LRs show how much more likely the patient is to get a *positive* or *negative* test if they had the disease, compared with a person without disease: LR+ is usually  $>1$  and LR- is usually  $<1$  [115].

There were no missing data for G8 items or GA findings. Few data were missing for

the 14 additional items: their proportion ranged from 0% to 4.9% (chronic renal failure) in the training set and 4.1% (health perception status) in the validation set. We imputed missing values using 10-fold multiple imputation by chained equations (MICE) and combining the estimates using Rubin's rules [116]. To specify, the procedure of multiple imputation (MI) uses the distribution of the observed data to estimate a set of likely values of the missing data. MI estimates these values  $m$  times, each time incorporating a random component to reflect the uncertainty about the missing values. After running the procedure,  $m$  different datasets are created on which the desired analysis is performed (e.g. logistic regression in our study). The MICE approach imputes data on a variable by variable basis by specifying an imputation model per variable, conditional on the other predictors and on the outcome. MICE operates under the assumption that given the variables used in the imputation procedure, the missing data are missing at random (MAR), which means that the probability that a value is missing depends only on observed values. In the final step, the  $m$  estimates are combined into an overall unbiased parameter estimate for each parameter in the model.

This observational study is reported according to the STARD checklist for diagnostic accuracy studies.

## 3.3 Results

### 3.3.1 Study population

Between January 2007 and October 2012, 1056 patients were included into the ELCAPA cohort (training set), of whom 729 had complete G8 data available at the time of our analysis. Between November 2012 and July 2014, 442 patients were included (validation set), of whom 414 had complete G8 data (**Figure 2.3**). **Table 3.1** reports the

general characteristics of the study populations. Overall, 632 patients (86.7%) had at least one impaired GA test in the training set (14.8% had 1 impaired test, 30.1% had 2 or 3 impaired tests, and 41.8% had 4 or more impaired tests) and 390 (94.2%) had at least one impaired GA test in the validation set (14.5% had 1 impaired test, 35.3% had 2 or 3 impaired tests, and 44.4% had 4 or more impaired tests).

Compared to the population not included in the development analysis (n=327), a higher proportion of patients had a better performance status (PS 0-1: 53.6% vs. 36.0%,  $p<0.0001$ ), had breast, prostate and urinary tract cancers and fewer patients had digestive and hematological cancers ( $p<0.0001$ ). More details in Annex 3, **Table A3.1**.

### 3.3.2 Univariate Analysis

**Table 3.2** reports the main results for the original G8 items and additional candidate items. Of the G8 items, 7 were significantly associated with an abnormal GA; the remaining item was BMI ( $p=0.06$ ). Anorexia, weight loss and prescription drugs were significantly associated with an abnormal GA in the 4 main cancer sites (Annex 4, **Table A4.1**). Of the 14 additional items, with the exception of gender ( $p >0.10$ ), 13 were significantly associated with an abnormal GA. Asthenia and ECOG-PS were significantly associated with an abnormal GA in the 4 main cancer sites (Annex 4, **Table A4.2**).

Odds ratios from the logistic regression analyses, showing items that predict impairment of the GA, are presented in Annex 5 (**Table A5.1**).

ROC analysis of the number of medications per day showed an optimal cutoff of  $\geq 6$ , as shown by the maximized Youden's index (33.0 [ $\geq 6$ ] vs. 9.9 [ $>3$ ]) and minimal distance between the ROC curve and upper left corner (**Figure 3.1**).

**Table 3.1.** Patient characteristics in the ELCAPA-07 cohort study

Characteristics	Training set	Validation set
	N=729	N=414
	N (%)	N (%)
Age in years, median (IQR)	80 (76;84)	81 (78;86)
Male gender	387 (53.1)	197 (47.6)
Living alone at home	274 (37.6)	156 (37.7)
Cancer site		
Breast	137 (18.7)	72 (17.4)
Colorectal	131 (17.9)	73 (17.6)
Upper digestive tract and liver	117 (16.0)	71 (17.1)
Urinary tract	118 (16.1)	61 (14.7)
Prostate	99 (13.5)	32 (7.7)
Hematological	49 (6.7)	25 (6.0)
Other <sup>a</sup>	78 (10.7)	79 (19.1)
Metastasis	299 (41.0)	141 (34.1)
Number of medications per day, median (IQR)	6 (4;8)	6 (4;9)
ECOG-PS		
0 - Fully active	177 (24.3)	58 (14.0)
1 - Restricted activity but ambulatory	213 (29.2)	128 (30.9)
2 - Up >50% of waking hours	129 (17.7)	82 (19.8)
3 - Confined to bed >50% of the day	148 (20.3)	85 (20.5)
4 - Completely disabled	61 (8.4)	60 (14.5)
Abnormal <sup>b</sup> geriatric assessment	632 (86.7)	390 (94.2)
ADL≤5	218 (29.9)	147 (35.5)
IADL≤7	457 (62.7)	282 (68.1)
MMSE≤23	193 (26.5)	124 (30.0)
Mini-GDS≥1	250 (34.3)	125 (30.2)
MNA≤23.5	426 (58.4)	256 (61.8)
CIRS-G≥1 comorbidity grade 3/4	414 (56.8)	282 (68.1)
TUG≥20 s	304 (41.7)	113 (27.3)

Abbreviations: IQR, interquartile range; ECOG-PS, Eastern Cooperative Oncology Group-Performance Status; ADL, Activities of Daily Living; IADL, Instrumental ADL; MMSE, Mini Mental State Examination; Mini-GDS, Mini Geriatric Depression Scale; MNA, Mini Nutritional Assessment; CIRS-G, Cumulative Illness Rating Scale for Geriatrics; TUG, Timed Up-and-Go test

<sup>a</sup> Training/Validation set: unknown primary origin (n=21/9), lung (n=17/22), skin (n=14/27), sarcoma (n=9/4), brain (n=5/6), gynecologic (n=4/5), others (n=8/6)

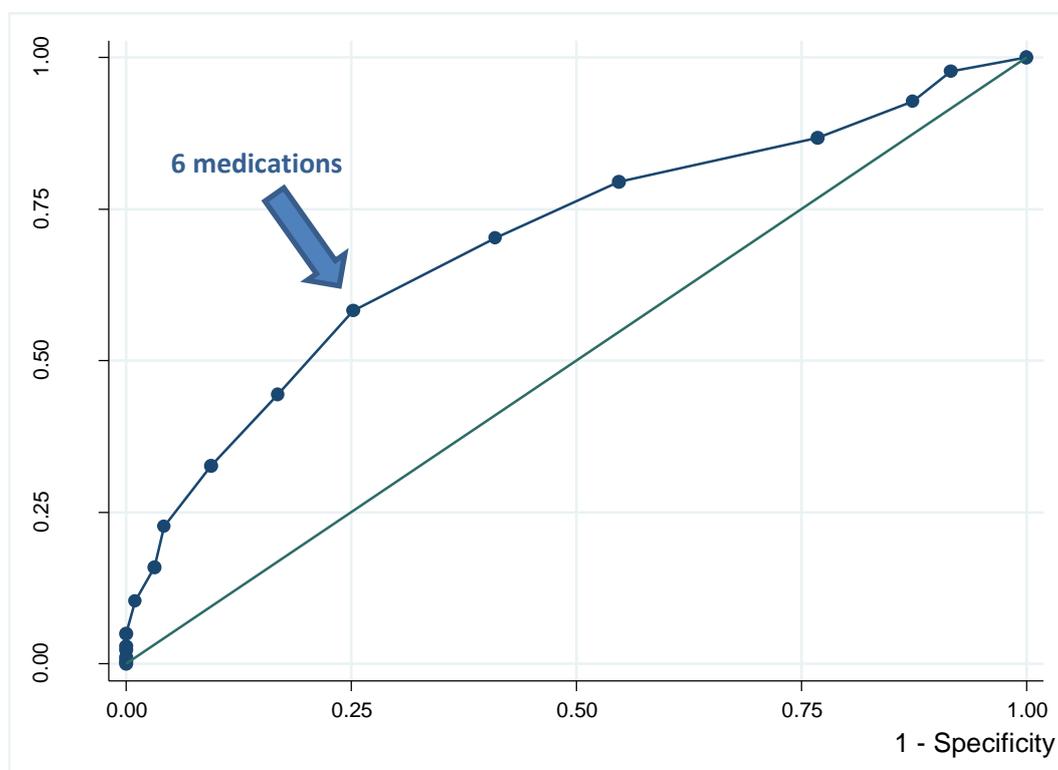
<sup>b</sup> ≥1 impaired score in ADL, IADL, MMSE, mini-GDS, MNA, CIRS-G and/or TUG

**Table 3.2.** Univariate analysis of the ability of candidate items to predict impairment of the geriatric assessment (training set; N=729)

	Normal GA (N= 97)	Abnormal GA (N=632)	P-value
<b>ORIGINAL G8 ITEMS</b>			
<b>Anorexia</b>			
Absent	88 (90.7%)	296 (46.8%)	<b>&lt;0.001</b>
Moderate	8 (8.2%)	232 (36.7%)	
Severe	1 (1.0%)	104 (16.5%)	
<b>Weight loss</b>			
Absent	80 (82.5%)	223 (35.3%)	<b>&lt;0.001</b>
1-3 kg	14 (14.4%)	119 (18.8%)	
Does not know	1 (1.0%)	63 (10.0%)	
>3 kg	2 (2.1%)	227 (35.9%)	
<b>Body Mass Index - BMI (kg/m<sup>2</sup>)</b>			
≥23	79 (81.4%)	425 (67.2%)	0.063
21≤ BMI< 23	7 (7.2%)	97 (15.3%)	
19≤ BMI< 21	8 (8.2%)	67 (10.6%)	
<19	3 (3.1%)	43 (6.8%)	
<b>Mobility</b>			
Goes out	97 (100%)	409 (64.7%)	<b>&lt;0.001</b>
Able to get out of bed/chair but does not go out	0 (0.0%)	131 (20.7%)	
Bed or chair bound	0 (0.0%)	92 (14.6%)	
<b>Dementia/Depression</b>			
Absent	84 (86.6%)	325 (51.4%)	<b>&lt;0.001</b>
Moderate	10 (10.3%)	245 (38.8%)	
Severe	3 (3.1%)	62 (9.8%)	
<b>Prescription drugs (&gt;3)</b>	52 (53.6%)	499 (79.0%)	<b>&lt;0.001</b>
<b>Self-rated health status</b>			
Better	58 (59.8%)	174 (27.5%)	<b>&lt;0.001</b>
As good	30 (30.9%)	207 (32.8%)	
Does not know	8 (8.2%)	167 (26.4%)	
Not as good	1 (1.0%)	84 (13.3%)	
<b>Age</b>			
<80 years	58 (59.8%)	289 (45.7%)	<b>0.024</b>
80-85 years	30 (30.9%)	227 (35.9%)	
>85 years	9 (9.3%)	116 (18.4%)	
<b>NEW ITEMS</b>			
<b>Gender, male</b>	44 (45.4%)	343 (54.3%)	0.102
<b>Living alone at home</b>	50 (51.6%)	224 (35.5%)	<b>0.002</b>
<b>Asthenia</b>	45 (46.4%)	487 (77.1%)	<b>&lt;0.001</b>

<b>Risk of fall (single-leg stance &lt;5 seconds)</b>	26 (26.8%)	435 (68.8%)	<b>&lt;0.001</b>
<b>Fall(s) in the 6 past months</b>	14 (14.4%)	209 (33.9%)	<b>&lt;0.001</b>
<b>ECOG-PS</b>			
0	71 (73.2%)	106 (16.8%)	<b>&lt;0.001</b>
1	24 (24.7%)	189 (29.9%)	
2/3/4	2 (2.1%)	336 (53.2%)	
<b>Metastasis</b>	32 (36.8%)	267 (50.3%)	<b>0.020</b>
<b>Urinary and/or fecal incontinence</b>	6 (6.2%)	159 (25.2%)	<b>&lt;0.001</b>
<b>Heart failure and/or Coronary Heart Disease</b>	6 (6.2%)	198 (31.3%)	<b>&lt;0.001</b>
<b>Complete Arrhythmia with Atrial Fibrillation</b>	6 (6.2%)	126 (19.9%)	<b>0.002</b>
<b>Heart rhythm disorder</b>	21 (21.7%)	102 (16.2%)	0.186
<b>Hypertension</b>	53 (54.6%)	431 (68.2%)	<b>0.008</b>
<b>Diabetes</b>	10 (10.3%)	160 (25.3%)	<b>0.002</b>
<b>Chronic renal failure</b>	37 (38.1%)	376 (59.5%)	<b>&lt;0.001</b>
<b>Chronic respiratory failure</b>	1 (1.0%)	49 (7.8%)	<b>0.042</b>

Abbreviations: GA, geriatric assessment; ECOG-PS, Eastern Cooperative Oncology Group-Performance Status: 0, fully active; 1, restricted activity but ambulatory; 2, up >50% of waking hours; 3, confined to bed >50% of the day; 4, completely disabled.



**Figure 3.1.** Receiver-operating characteristic (ROC) curve of number of medications / day for predicting impairment of the geriatric assessment

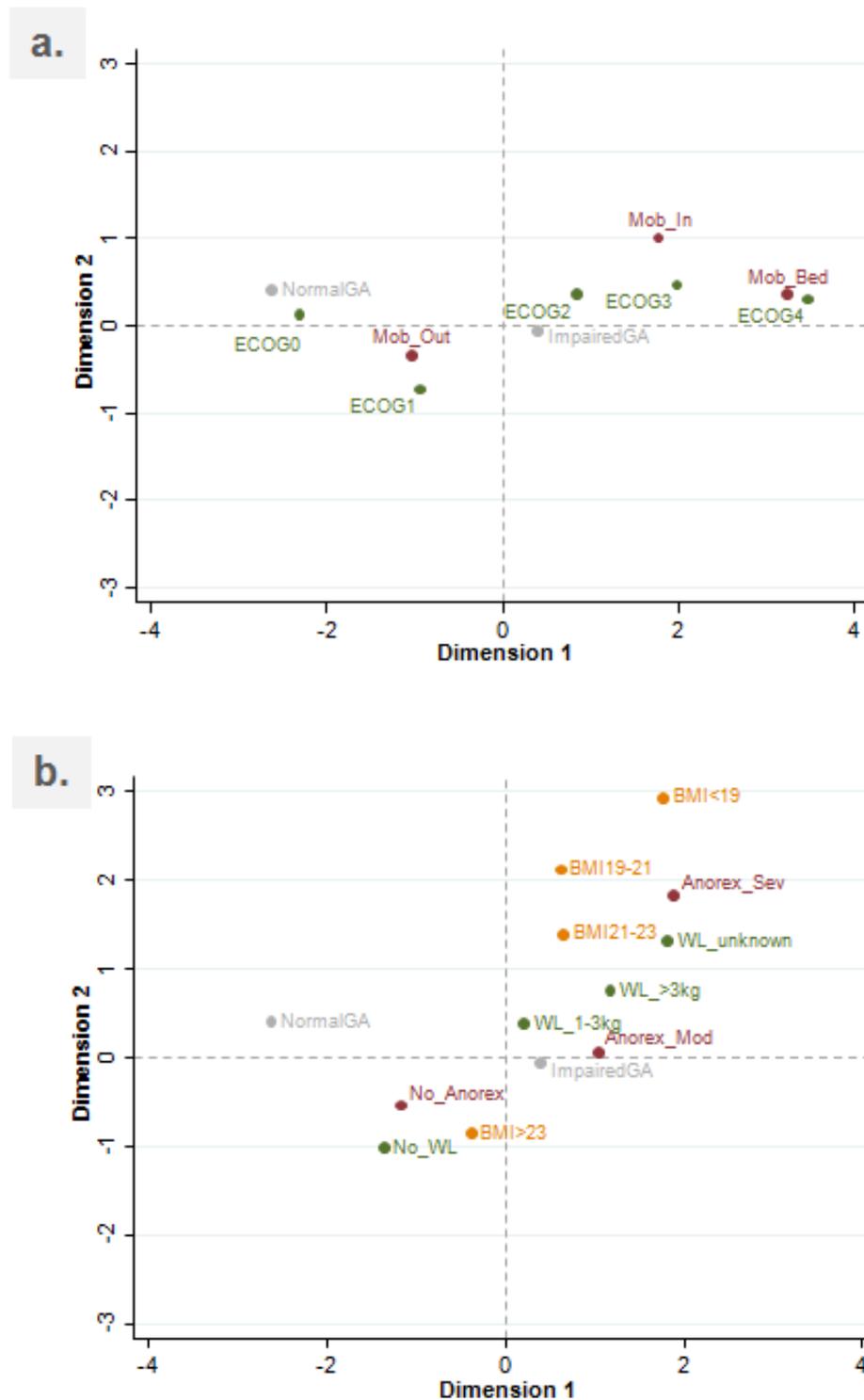
### 3.3.3 Multiple Correspondence Analysis

MCA was conducted to assess the correlations between items. All G8 items (prescription drugs with the optimal cutoff of  $\geq 6$ ) and the new candidate items associated with an abnormal GA at the  $P < 0.05$  level, were entered into the MCA as active variables (contributing to the identification of factors), whereas an (ab)normal GA was entered as an illustrative variable. Two-dimensional plots of factor scores were created to visualize interrelationships between variable modalities (more details in **Annex 6**).

**Figure 3.2** shows the contributions of conceptually similar items to the first factor (horizontal axis; Dimension 1) plotted against the second factor (vertical axis; Dimension 2) for the ECOG-PS and G8 mobility items (a), G8 weight loss, BMI, and anorexia items (b), risk of falls (assessed by the single-leg stance time) and history of falls in the past 6 months (c), and comorbidities, age, and medications (d). The first factor (Dimension 1) was associated with the presence (or absence) of items related to an abnormal GA, with category points located in the negative part (left) being associated with no impairment and those in the positive part (right) indicated increasing GA impairment. Categories from ECOG-PS, asthenia, mobility, nutrition, and fall-related items were well represented on this axis. The second factor (Dimension 2) was related to the presence or absence of comorbidities, number of medications per day, and age.

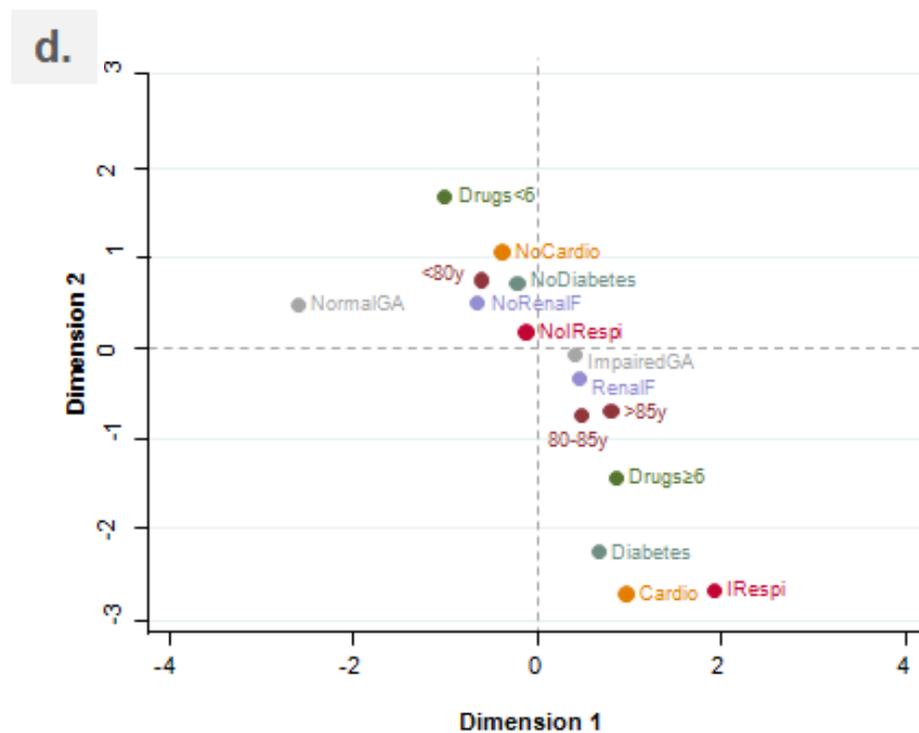
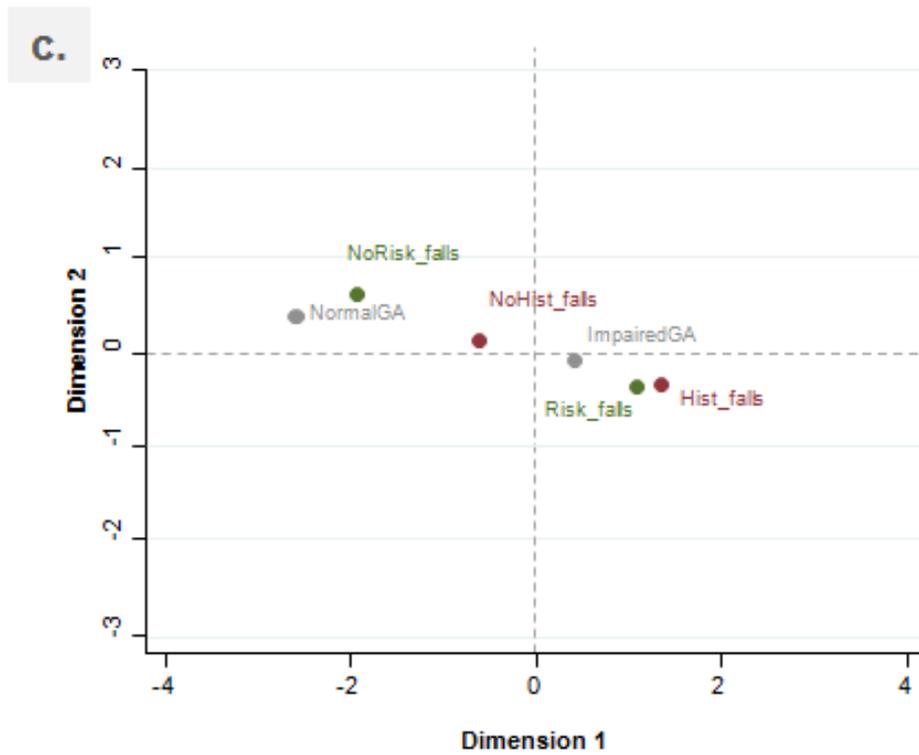
Because several variables were located in close proximity, only those items exhibiting the highest discriminative power and showing a graded distribution along the first MCA axis were kept for the multivariate analysis. These items were the ECOG-PS, weight loss, and fall risk. Thus, we omitted the G8 BMI, anorexia, mobility, and history of fall in the past 6 months.

Figure 3.2. Multiple correspondence analysis of candidate items



**Legend**

Abbreviations: Normal GA/Abnormal GA, normal/abnormal geriatric assessment; ECOG-PS, Eastern Cooperative Oncology Group - Performance Status; Mob\_Out/Mob\_In/Mob\_Bed, G8 mobility item (goes out, gets out of bed/chair, bed/chair bound); BMI, body mass index; WL, weight loss; No\_Anorex/Anorex\_Mod/Anorex\_Sev, G8 decreased food intake/anorexia item (no, moderate, severe).



**Legend**

Abbreviations: Normal GA/Abnormal GA, normal/abnormal geriatric assessment; (No)Risk\_falls, risk of falls, single-leg stance <5 s (≥5 s); (No)Hist\_falls, history of falls in the past 6 months; (No)Cardio, heart failure/coronary heart disease; (No)IRespi, chronic respiratory failure; (No)RenalF, chronic renal failure; <80y/80-85y/85y, age in years.

### 3.3.4 Multivariate Model

Based on the previous steps, we selected 14 items for the multivariate analysis. A backward stepwise approach showed that six items were independently associated with an abnormal GA. We merged several categories when similar OR values were found across adjacent modalities or for certain modalities that almost perfectly predicted an abnormal GA (ECOG-PS, weight loss, dementia/depression, and self-rated health status), yielding the final model shown in **Table 3.3**. Model calibration was excellent ( $\chi^2 = 69.8$ ;  $p=0.97$ ). The regression coefficients were then rescaled and rounded to integers to provide weights suitable for use in clinical practice. Among tested multiplication coefficients, 3.5 proved optimal. The final six-item questionnaire is shown in **Table 3.4**.

### 3.3.5 Diagnostic Performance of the Original versus Modified G8

Using the recommended score cutoff of  $\leq 14$  (11), the original G8 demonstrated the following indices: sensitivity, 87.2% (95% CI, 84.3–89.7); specificity, 57.7% (47.3–67.7); PPV, 93.1% (90.7–95.0); NPV, 40.9% (32.6–49.6); LR+, 2.06 (1.63–2.61); and LR-, 0.22 (0.17–0.29). For the modified G8, the cutoff of  $\geq 6$  of 35 points maximized sensitivity and yielded the following characteristics: sensitivity, 89.2% (86.5–91.5); specificity, 79.0% (69.4–86.6); PPV, 96.5% (94.7–97.9); NPV, 52.8% (44.3–61.2); LR+, 4.24 (2.87–6.26); and LR-, 0.14 (0.11–0.18). Using the higher cutoff of  $\geq 7$  of 35 produced the following values: sensitivity, 85.8% (82.8–88.5); specificity, 88.4% (80.2–94.1); PPV, 98.0% (96.4–99.0); NPV, 48.8% (41.2–56.6); LR+, 7.41 (4.25–12.93); and LR-, 0.16 (0.13–0.20). The AUROC was 86.5% (83.5–89.6) for the original G8 and 91.6% (89.3; 93.9) for the modified G8 ( $p=0.0002$ ) (**Figure 3.3**). When we analyzed each cancer site, we found that the modified G8 yielded consistently higher AUROC values with greater uniformity

(from 87.1% [colorectal cancers] to 96.2% [urinary tract cancers]) compared with the original G8 (from 78.4% [colorectal cancer] to 93.6% [liver/upper gastrointestinal tract cancer]) (Figure 3.4).

### 3.3.6 Internal and Temporal Validation and Sensitivity Analyses

The internal validation procedure showed no evidence of overoptimism (optimism = 0.89% ± 0.13%). The bias-corrected AUROC was 90.7% for the final prediction model. Sensitivity analyses consisted first in including a social domain in the GA, evaluated by an inadequate social environment (absence of a primary caregiver or adequate support at home or a strong circle of family and friends able to meet the needs of the patient at the time of the evaluation), which yielded an AUROC of 90.6% (95% CI, 88.1-93.2), with a sensitivity of 87.4% (84.6-89.9) and a specificity of 77.5% (66.8-86.1) with a cutoff for impairment of ≥1 test among the 8 tests. Other sensitivity analyses found an AUROC of 91.4% (89.2–93.6) when the CIRS-G was removed from the GA, 90.4% (88.0–92.8) when the ADL/IADL was removed from the GA, 90.3% (88.0–92.6) when an abnormal GA was defined as impairment of ≥2 tests among the 7 tests, and 89.8% (87.4-92.2) among the 8 tests (including the social domain).

Applying the modified G8 to the temporal validation set produced the following values: AUROC, 92.8% (88.4–97.2); at the ≥6 of 35 cutoff: sensitivity, 91.4% (88.0–94.1); specificity, 75.0% (53.3–90.2); PPV, 98.2% (96.0–99.3); NPV, 37.5% (24.0–52.6); at the ≥7 of 35 cutoff: sensitivity, 88.0% (84.1–91.2); specificity, 87.5% (67.6–97.3); PPV, 99.0% (97.2–99.8); and NPV, 33.3% (22.0–46.3). Model calibration was excellent ( $\chi^2 = 3.12$ ;  $p=0.93$ ).

**Table 3.3.** Final multivariate logistic model for predicting impairment of the geriatric assessment (training set; N=729)

	Regression				P-value	Final weights <sup>a</sup>
	Coefficient		Odds Ratio			
	Estimate	95% CI	Estimate	95% CI		
No weight loss	0 (ref)	-	1 (ref)	-	-	<b>0</b>
Weight loss 1-3 kg	0.70	(0.00;1.40)	2.01	(1.00;4.06)	0.052	<b>2</b>
Weight loss >3 kg / unknown	2.77	(1.56;3.97)	15.90	(4.75;53.23)	<0.0001	<b>10</b>
Dementia / Depression	0.84	(0.13;1.55)	2.32	(1.14;4.73)	0.020	<b>3</b>
Drugs/day ≥ 6	0.64	(0.04;1.23)	1.89	(1.04;3.43)	0.036	<b>2</b>
Lower self-rated health status	0.87	(0.06;1.69)	2.40	(1.06;5.43)	0.036	<b>3</b>
ECOG-PS grade 0	0 (ref)	-	1 (ref)	-	-	<b>0</b>
ECOG-PS grade 1	1.15	(0.57;1.73)	3.16	(1.77;5.65)	<0.0001	<b>4</b>
ECOG-PS grade 2, 3, or 4	3.31	(1.84;4.78)	27.39	(6.28;119.39)	<0.0001	<b>12</b>
Heart failure and/or coronary heart disease	1.35	(0.42;2.27)	3.85	(1.53;9.71)	0.004	<b>5</b>

Abbreviations: 95% CI, 95% confidence interval; ECOG-PS, Eastern Cooperative Oncology Group-Performance Status.

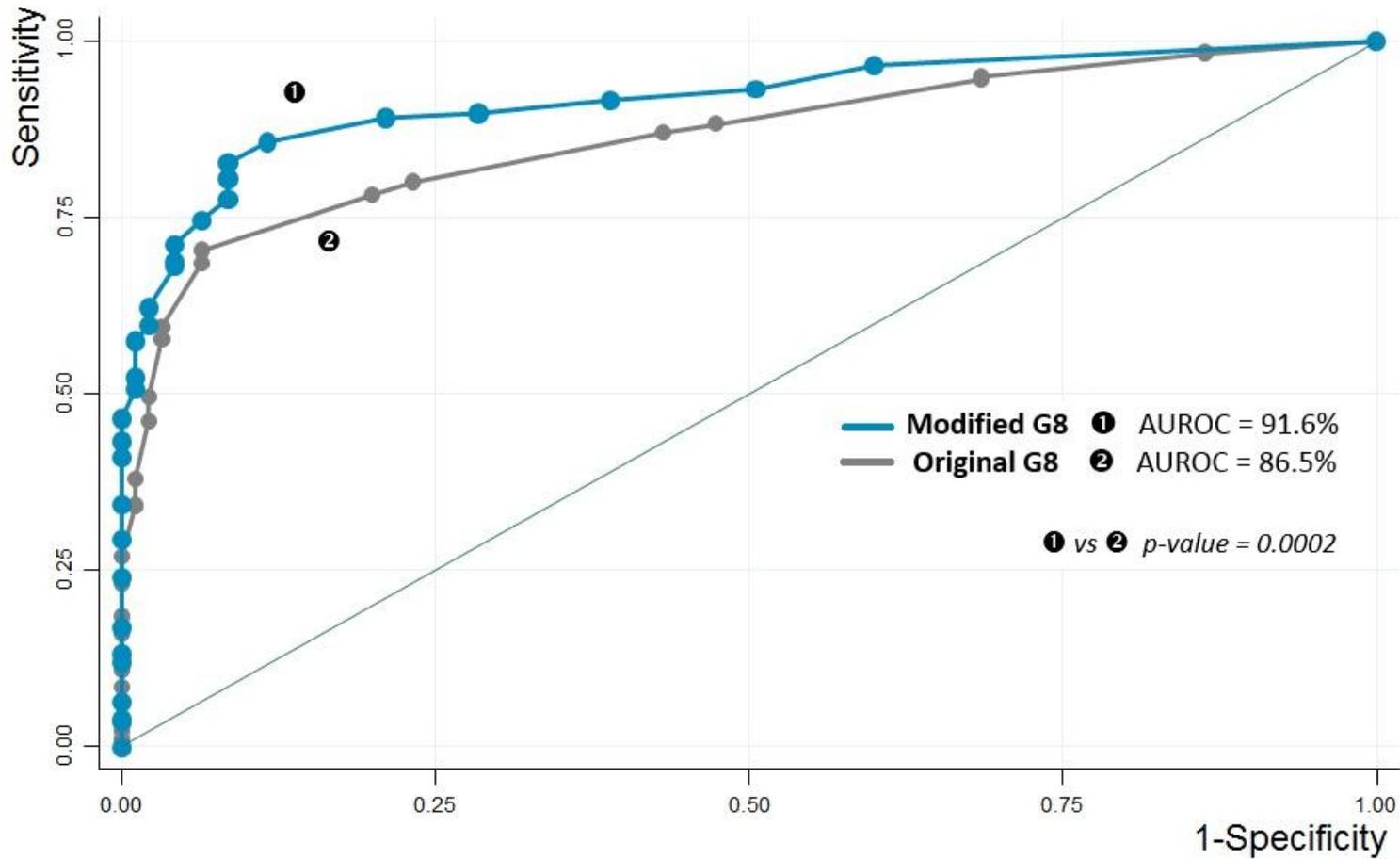
<sup>a</sup> Calculated by multiplying the regression coefficients by 3.5 then rounding the results to integers.

**Table 3.4.** Final modified G8: the ELCAPA-07 cohort study

<i>Items</i>	<i>Score</i>
<b>1</b> Weight loss during the past 3 months	
>3 kg / patient does not know	<b>10</b>
1-3 kg	<b>2</b>
No weight loss	<b>0</b>
<b>2</b> Neuropsychological problems	
Mild / severe dementia or depression	<b>3</b>
No neuropsychological problems	<b>0</b>
<b>3</b> Takes at least six drugs per day	
Yes	<b>2</b>
No	<b>0</b>
<b>4</b> Compared to other people of the same age, how does the patient rate his or her health status?	
Not as good / patient does not know	<b>3</b>
As good or better	<b>0</b>
<b>5</b> Performance Status (PS)	
PS 2, 3, or 4: Ambulatory but unable to carry out any work activities / Confined to bed >50% / Disabled	<b>12</b>
PS 1: Restricted in physically strenuous activity but ambulatory and able to carry out work of a light or sedentary nature	<b>4</b>
PS 0: Fully active	<b>0</b>
<b>6</b> Past history of heart failure or coronary artery disease	
Yes	<b>5</b>
No	<b>0</b>
<b>Total</b>	<b>___ / 35</b>

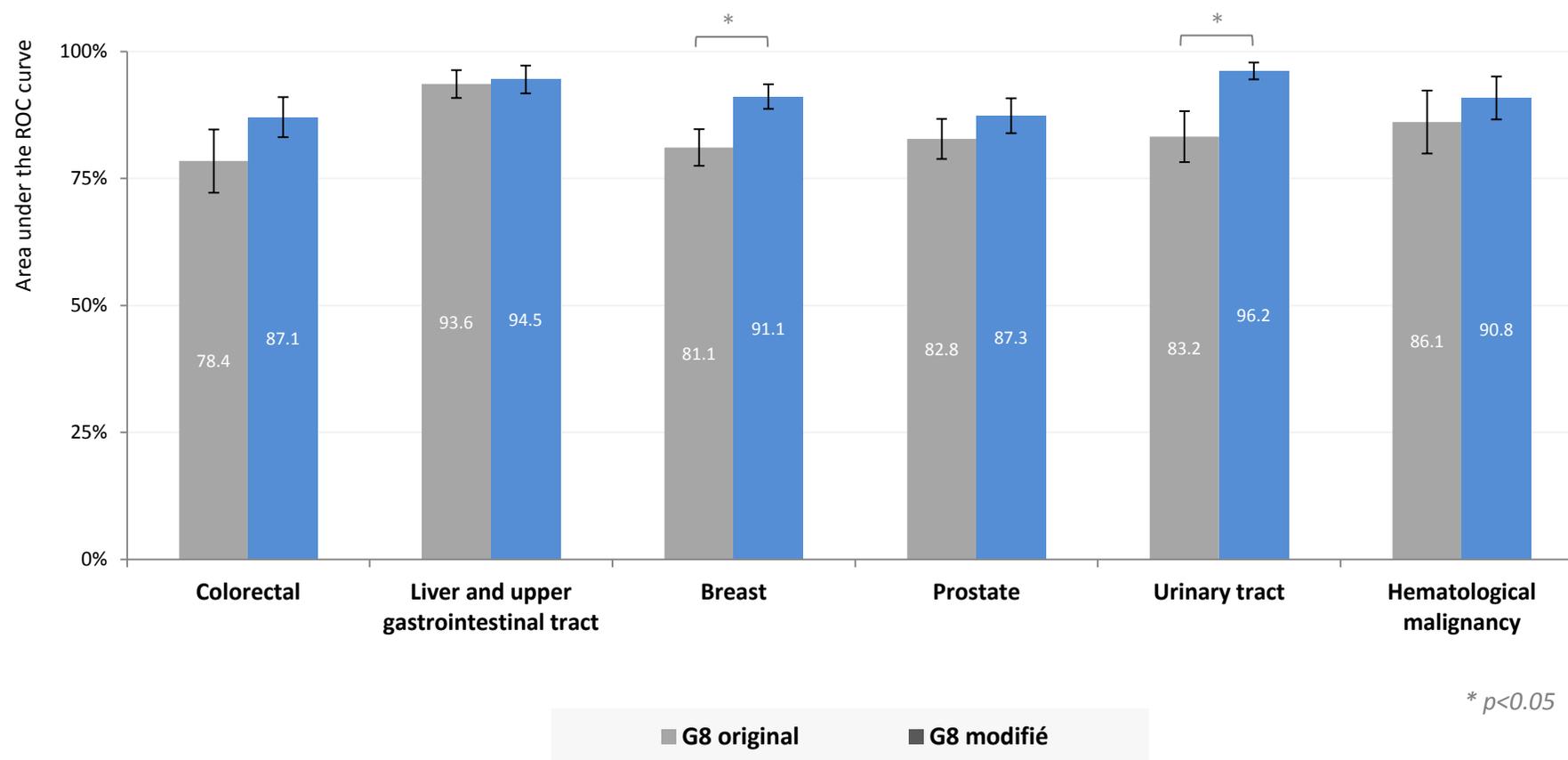
The total score is the sum of the scores on each of the 6 items.

A total score  $\geq 6$  is considered abnormal and warrants a full geriatric assessment in the two-step approach.



**Figure 3.3.** ROC curves for predicting impairment of the geriatric assessment, used as the reference test: original versus modified G8 questionnaire (training set; n = 729)

Abbreviations: AUROC, area under the ROC curve; ROC, receiver-operating characteristic.



**Figure 3.4.** Area under the ROC curve by cancer site: original versus modified G8

Abbreviation: ROC, receiver-operating characteristic.

## 3.4 Discussion

### 3.4.1 Summary of findings

We developed a modified version of the G8 based on six simple items that are routinely collected by geriatricians. The modified G8 demonstrated a sensitivity of 89.2% and a specificity of 79.0% at the optimized cutoff of  $\geq 6$  of 35 points, with evidence of homogenous performance across tumor sites.

Our objective was to develop a variant of the G8 that would improve the identification of patients requiring a full GA. We used a systematic approach, as typically applied for developing clinical rules or prediction models. This approach consisted of the initial selection of candidate items based on clinical reasoning, multivariate analyses to identify items conveying independent information, internal validation based on bootstrapping techniques to prevent overfitting [110], and reassessment of the model performance on a validation set population. Special attention was given to weights computation to obtain an easy-to-use tool while limiting the loss of information inevitably associated with rounding [113]. Finally, we used multiple imputation at each step of model development to maintain an effective sample size and to control their potential influence on the final model [110].

The modified G8 has only six items yet covers multiple domains included in the GA, namely, nutritional status, mood or cognition, comorbidities, and polypharmacy, in addition to self-rated health status and a simplified version of the ECOG-PS. Interestingly, alterations in these items have been shown to predict adverse outcomes [67,94,117,118]. For the number of medications per day, instead of the  $>3$  cutoff used in the G8, we found that the  $\geq 6$  cutoff improved discrimination, in keeping with

conclusions from a recent expert consensus conference [119]. Similarly to the original G8, most items in the modified G8 are subjective and therefore not well suited as criteria for individual diagnosis [61]. The assessment of a past history of heart failure/CHD should not be viewed as an abbreviated version of comorbidity assessment tools such as the CIRS-G but rather as a marker predicting an abnormal GA when used in combination with the other items. Despite being associated with an abnormal GA in univariate analysis, several items (age, BMI, anorexia, and mobility) were omitted from the modified G8 because they had minimal independent discriminative power, given their close correlations with other variables, as visualized by MCA.

#### 3.4.2 Comparison with other reviews

A wide variety of screening tools have been evaluated to identify patients likely to benefit from a complete GA. The G8 [61], the Vulnerable Elders Survey-13 [79], and the Groningen Frailty Indicator [120] are among the most extensively assessed. Although none of the available tools is markedly better than the others [56], the G8 has the theoretical advantage of having been specifically developed for older patients with cancer, with a selection of items covering important domains in this population [61]. A recent review identified eight studies evaluating the ability of the G8 to predict an abnormal GA [56]. Sensitivity was usually high, with a range of 65%–92% (median, 85.5%), but specificity was lower, ranging from 3% to 75% (median, 59.5%). Similarly, in our study, the original G8 was 87.2% sensitive but only 57.7% specific. It is noticeable that the G8 was derived from the MNA-SF questionnaire, because of its known high prognostic value for survival in older patients [61]. The fact that the MNA-SF was not designed to specifically detect an abnormal GA probably explains the lack of specificity of the G8 as a screening instrument.

### 3.4.3 Strengths and limitations

The modified G8 was robust to sensitivity analyses involving changes in the definition of GA/abnormal GA. This point is of particular interest, because various definitions have been used in previous studies assessing the performance of the G8 in the absence of a clearly defined reference GA [56,57]. Moreover, the modified G8 showed homogeneity across tumor sites, including various solid tumors and hematological malignancies, whereas evidence for heterogeneity was previously reported for the original G8 [67,71,121].

Should the improved screening performance of the modified G8 be confirmed, this new tool may encourage the actual use of a two-step approach, in which the results of screening determine whether a full GA is performed [56]. High discriminative power is essential to avoid performing time-consuming unnecessary GAs (false positives) and to ensure that no patients requiring a GA are missed (false negatives). This last point is of major importance, given the consistently high prevalence of abnormal GA findings in several studies conducted in various settings (>80% (11, 12, 24); 86.7% in our study).

This study has several limitations. First, in keeping with our study objective, we confined our sample to patients for whom the GA and G8 items were available, which resulted in 327 patients being excluded from the original sample of 1,056 patients. However, we found no statistically significant differences between included and excluded patients regarding the main demographic and clinical features (age, gender, cancer type, and cancer spread; Annex 3, **Table A3.1**), suggesting minimal selection bias. Second, several potentially relevant variables were not entered in our database at the time of the analysis, including specific items from validated scales or details on the social environment. These variables deserve investigation in future studies.

### **3.5 Conclusion**

Based on a large prospective cohort study, we developed a modified G8 screening tool that exhibited better diagnostic performance across a variety of tumor sites and greater parsimony, with only six items instead of eight (original G8), facilitating selection for a complete geriatric assessment. Our work illustrates the usefulness of combining in-depth statistical analyses with expert judgment to ensure both optimal discriminative power and clinical relevance.

Further research is needed to confirm the features of the modified G8 in other populations and to measure its prognostic value and its impact on treatment decisions and health outcomes.

---

This work was published in *the Oncologist* (Martinez-Tapia et al., 2015)



## CHAPTER 4

---

### **External validation Analysis**

## 4.1 Introduction

The 6-item modified G8 screening tool for older patients with cancer was developed in the ELCAPA-07 Prospective Cohort Study [122] as a primary objective of this thesis. It presented better diagnostic performance and greater uniformity across cancer sites. This modified tool may permit a better selection of older patients with cancer for a full GA. However, model's good diagnostic performance in the development sample is not sufficient to confirm that a model is valuable, even when complemented with internal validation techniques to provide estimates corrected for overfitting and optimism [123]. Indeed, the performance of prediction models is generally better on the data set on which the model has been developed. It is therefore essential to evaluate the predictive performance of the model in datasets that were not used for its development and preferably selected from different settings, before implementation in practice [124]. These so called external validation studies provide estimates of a model's accuracy in new populations, assess the agreement between predicted and observed risks and thus test the generalizability of a model. If validity indices are deemed insufficient, a subsequent updating can be performed. Updating methods adjust the model to new circumstances or settings to optimize its performance. Several methods for updating prediction models have been proposed in the literature and evaluated in different contexts [110,123,125-129]. They vary in extensiveness (i.e. number of adjusted or re-estimated parameters) ranging from simple adjustment of the baseline risk (intercept) to additional adjustment of predictors weights or even adding or removing predictors. **Table 4.1** summarizes the updating methods and their principles for applicability.

**Table 4.1.** Updating methods for predictions models

Method	Updating method	Reason for updating
0	No adjustment (the original prediction model)	-
1	Adjustment of the intercept (baseline risk)	Difference in the outcome frequency (prevalence or incidence) between development and validation sample
2	Method 1 + adjustment of all predictor regression coefficients by one overall adjustment factor	Regression coefficients of the original model are overfitted (or underfitted)
3	Method 2 + extra adjustment of regression coefficients for predictors with different strength in the validation sample as compared with the development sample	As in method 2, and the strength (regression coefficient) of one or more predictors may be different in the validation sample
4	Method 2 + stepwise selection of additional predictors	As in method 2, and one or more potential predictors were not included in the original model, or a newly discovered marker may need to be added
5	Re-estimation of all regression coefficients, using the data of the validation sample only	The strength of all predictors may be different in the validation sample, or the validation sample is much larger than the development sample
6	Method 5 + stepwise selection of additional predictors	As in method 5, and one or more potential predictors were not included in the original model

Source: Moons et al. [123]

The aim of this study was to externally validate the diagnostic performance of the 6-item modified G8 score for discriminating between patients with normal and abnormal GA using data from a multicenter prospective cohort of older patients with cancer. We further investigated whether updating methods may improve the performance of the modified G8. Finally, and because this cohort originally served as the external validation study for the original G8 score, we also compared the diagnostic performances of the modified G8 to that of the original G8.

## 4.2 Methods

### 4.2.1 Study population

The ONCODAGE prospective cohort is described in Chapter 2 (General methodology). **Figure 2.3** presents the flow chart of participants included in this study.

### 4.2.2 Reference standard

As in the development cohort, the multidimensional GA was used as the reference standard to evaluate the performance of the modified G8 in the independent cohort. An abnormal GA was defined as at least one impaired component among seven validated tests previously mentioned [122]. It should be noted that the Geriatric Depression Scale-15 items (GDS-15) replaced the mini-GDS used in the development study, as it wasn't available in the validation dataset.

### 4.2.3 Screening tools

Each screening tool has been described previously (**Table 1.6** [original G8] and **Table 3.4** [modified G8]). It should be noted that since information about the item “past history of heart failure or coronary heart disease” of the modified G8 was not explicit in the external validation dataset, we used the CIRS-G heart category as a substitute, assuming history of heart disease if severity grade  $\geq 1$ .

### 4.2.4 Statistical analysis

For each patient in the validation cohort, the individual score and predicted probabilities were calculated using the prognostic factors and respective integer weights from the regression coefficients as estimated in the development data cohort.

Descriptive analyses (details in section 2.4) were used to compare populations in terms of demographic and clinical characteristics, as well as outcome and predictors.

We used the framework for external validation from Debray et al. [130]. Therefore, we first assessed the extent to which the development and validation samples had a similar case-mix, by comparing the mean linear predictor (LP) and standard deviation (SD) of the model in each cohort. Differences in mean of the LP between the development and validation samples reflect the difference in predicted outcome frequency (revealing the calibration-in-the-large), and differences in the SD reflect the heterogeneity of case mix between the samples. Predictor effects in both cohorts were also evaluated.

The predictive performance of the modified G8 was then assessed by examining measures of calibration and discrimination [131].

*Calibration* is the agreement between the predicted and observed probabilities of an abnormal GA. Model calibration was first examined using the Hosmer-Lemeshow (HL) Goodness-of-Fit test and two other alternatives that do not require grouping of the data: Standardized Pearson's chi-square and Stukel's test [132]. A non-significant HL test indicates good calibration, whereas for the 2 alternative tests, a non-significant value indicates good fit. The observed frequencies versus the predicted probabilities for each tenth of predicted risk of the outcome were plotted (calibration plot). A smooth, nonparametric calibration line was added using a locally weighted scatter plot smoother (i.e., the loess algorithm) allowing us to examine calibration across the entire range of predicted probabilities. The calibration slope, visualized in the calibration plot, reflects whether the effects of the predictors in the validation samples are on average similar to the effects in the development sample and should lie on or near the diagonal reference line. It was calculated by estimating the regression coefficient in a logistic regression

model with the outcome (abnormal GA) as dependent variable and the LP of the model as the only covariate and is ideally 1. The calibration intercept (calibration-in-the-large) was further estimated by fitting a logistic regression model with the LP as an offset variable (setting the regression coefficient to 1). The intercept relates to calibration-in-the-large, which compares the mean of predictions with the mean outcome frequency. It therefore indicates whether predictions are in general correct and is ideally 0. Additionally, the average difference between predicted and observed abnormal GA was calculated (MAPE, mean absolute prediction error [133]). This measure is expressed by a number between 0 and 1, with values closer to 0 indicating better performance.

*Discrimination* is the ability of the model to differentiate between patients with or without an abnormal GA. This measure was quantified by calculating the AUROC (c-statistic); a value of 0.5 represents chance (poor discrimination) and 1 represents perfect discrimination [134]; values between 0.7 and 0.8 indicate good discrimination, values between 0.8 and 0.9, very good discrimination and greater than 0.9, excellent discrimination. Sensitivity, specificity, PPV, NPV, LR+ and LR- were additionally calculated.

The overall performance of the model was evaluated using the Brier score [135], comprising a number between 0 and 1, lower values indicating better performance, although a cutoff of <0.25 reflects the usefulness of a risk prediction model.

As a final step of the external validation framework, we combined the results of the 2 previous steps to interpret the performance of the modified G8, which suggested that an updating of the model was necessary.

Finally, the AUROC were compare between both tools (original and modified G8) in the whole included population and in the four main cancer sites.

## 4.3 Results

### 4.3.1 Study population

Between August 2008 and March 2010, 1435 patients were included into the ONCODAGE project, of whom 1304 had complete GA information and modified G8 data available (**Figure 2.3**). Compared to the non-included population (n=131), a lower proportion of included patients had a good performance status (PS 0-1: 75.1% vs. 93.8%,  $p < 0.0001$ ), and the proportion of men was higher (Annex 3, **Table A3.2**).

**Table 4.2** presents the main baseline characteristics including the outcome and predictors of the validation cohort compared with the development cohort.

Compared to the development cohort, patients in the validation cohort were younger, fewer patients presented metastases (17% vs. 41%) and the proportion of men was lower (31% vs. 53%) due to differences in cancer sites (i.e. breast cancer: 53% vs. 19%; prostate cancer: 9% vs. 14%). Other significant differences in distribution of characteristics between the 2 cohorts were observed for the fact of living alone and the number of medications per day.

The prevalence of abnormal GA was 79.4% in the validation cohort compared to 86.7% in the development cohort ( $p$ -value  $< 0.0001$ ). The distribution of most predictors varied significantly. Compared to the development cohort, patients in the validation cohort had less neuropsychological problems, took less medications, had better self-rated health status and better performance status (PS 0-1: 75% vs. 54%). Conversely, patients showed more frequently past history of heart failure or coronary artery disease (48% vs. 28%). The median score of the modified G8 also differed significantly between both cohorts: 11 (validation cohort) vs. 16 (development cohort).

**Table 4.2.** Distribution of patient's characteristics, outcome and predictors

Characteristics	N (%)		P-value
	Development Set (N=729)	Validation Set (N=1304)	
<b>Age, median (IQR)</b>	80 (76-84)	78 (74-82)	<0.0001
<b>Male gender</b>	387 (53.1)	406 (31.1)	<0.0001
<b>Living alone</b>	274 (37.6)	560 (42.9)	0.015
<b>Metastases</b>	299 (41.0)	183 (16.8)	<0.0001
<b>Cancer site</b>			<0.0001
Colorectal	131 (17.9)	191 (14.6)	
Upper gastrointestinal tract	117 (16.0)	0 (0.0)	
Prostate	99 (13.6)	112 (8.6)	
Urinary tract	118 (16.1)	0 (0.0)	
Breast	141 (19.3)	688 (52.8)	
Hematological malignancies	49 (6.7)	103 (7.9)	
Lung/Bronchial	17 (2.3)	140 (10.7)	
Others <sup>a</sup>	57 (7.8)	70 (5.4)	
<b>Abnormal <sup>b</sup> Geriatric Assessment Score modified G8</b>	632 (86.7)	1035 (79.4)	<0.0001
Abnormal ( $\geq 6$ )	588 (80.7)	934 (71.6)	<0.0001
Median (IQR)	16 (7-25)	11 (5-19)	<0.0001
<b>Weight loss during the past 3 months</b>			<0.0001
No weight loss	303 (41.6)	672 (51.5)	
1-3 kg	133 (18.2)	226 (17.3)	
>3 kg / patient does not know	293 (40.2)	406 (31.1)	
<b>Neuropsychological problems</b>	320 (43.9)	402 (30.8)	<0.0001
<b>Polypharmacy (<math>\geq 6</math> medications/day)</b>	386 (52.9)	505 (38.7)	<0.0001
<b>Lower self-rated health status</b>	260 (35.7)	383 (29.4)	0.003
<b>Performance Status</b>			<0.0001
PS 0: Fully active	177 (24.3)	523 (40.1)	
PS 1: Restricted in physically strenuous activity but ambulatory, able to carry out work of a light or sedentary nature	213 (29.2)	457 (35.1)	
PS 2, 3, or 4: Ambulatory but unable to carry out any work activities / Confined to bed >50% / Disabled	338 (46.4)	324 (24.9)	
<b>Past history of heart failure or coronary artery disease</b>	204 (28.0)	626 (48.0)	<0.0001

IQR, interquartile range

<sup>a</sup> Development/Validation cohort 2: unknown primary origin (n=20/0), skin (n=15/0), sarcoma (n=9/0), brain (n=5/0), upper aerodigestive tract (n=5/70), others (n=3/0).<sup>b</sup> Defined as an impaired score on  $\geq 1$  of the seven tests used in the geriatric assessment (ADL, IADL, MMSE, mini-GDS [Development dataset]/GDS [Validation dataset], MNA, CIRS-G, and TUG).

## 4.3.2 Investigating relatedness of development and validation samples

The regression coefficients of the prognostic factors in each sample are presented in **Table 4.3**. By refitting the original model in the validation sample, the heterogeneity in predictor-outcome associations between the development and validation samples was evaluated. The effects of the predictors were similar between samples, with the exception of the 3<sup>rd</sup> category of the item weight loss and the item past history of heart failure or coronary heart disease, which had decreased effects in the validation sample compared to the development sample.

Following the framework proposed by Debray et al. the mean and SD of the LP of both cohorts were also calculated. We found a mean difference of 1 for the LP (mean LP=3.2 [validation set] vs. 4.2 [development set]; p-value<0.0001) and a decreased spread (standard deviations=2.8 vs. 3.0).

**Table 4.3.** Coefficients of the logistic regression model (modified G8 score items) for predicting impairment of GA (development and validation sets)

Score <sup>a</sup>		Development dataset (N=729)		Validation dataset (N=1304)	
		Coefficient (95% CI)	P-value	Coefficient (95% CI)	P-value
0	No weight loss	0 (ref)		0 (ref)	
2	Weight loss 1-3 kg	0.69 (0.48;0.90)	<0.0001	0.76 (0.34;1.18)	<0.0001
10	Weight loss >3 kg / unknown	2.76 (2.40;3.12)	<0.0001	2.08 (1.46;2.71)	<0.0001
3	Dementia / Depression	0.28 (0.21;0.35)	<0.0001	0.40 (0.24;0.56)	<0.0001
2	Drugs/day ≥ 6	0.32 (0.23;0.41)	<0.0001	0.34 (0.15;0.54)	0.001
3	Lower self-rated health status	0.29 (0.21;0.38)	<0.0001	0.24 (0.08;0.39)	0.003
0	ECOG-PS grade 0	0 (ref)		0 (ref)	
4	ECOG-PS grade 1	1.16 (0.98;1.33)	<0.0001	1.02 (0.67;1.38)	<0.0001
12	ECOG-PS grade 2, 3, 4	3.30 (2.86;3.74)	<0.0001	2.88 (1.86;3.91)	<0.0001
5	Heart failure / CHD	0.27 (0.22;0.33)	<0.0001	0.08 (0.01;0.15)	0.021
	Constant	-0.61 (-0.74;-0.48)	<0.0001	-0.49 (-0.75;-0.24)	<0.0001

Abbreviations: 95% CI, 95% confidence interval; ECOG-PS, Eastern Cooperative Oncology Group-Performance Status; CHD, Coronary heart disease.

<sup>a</sup> Modified G8 final weights.

### 4.3.3 Assessment of the model's performance in the validation study

**Table 4.4** summarizes the results of the main measures evaluating the performance of the modified G8 in the validation set compared to the development set (overall performance, calibration and discrimination measures). Overall, the model had a good performance with a Brier score  $<0.25$ .

#### *Calibration*

Calibration was suboptimal in the validation cohort, shown by the calibration intercept of  $-0.25$  (82.1% predicted risk compared to 79.4% observed risk). There was a tendency for over-predicting risk with a calibration slope of 0.74. This was also shown in the calibration plot (**Figure 4.1(a)**), indicating poor agreement between the predicted and observed risks in some deciles of risks. In the low risk group ( $n=159$ ), the observed mean of abnormal GA was bigger to that predicted by the model; however, in higher risk groups (from 3<sup>rd</sup> to 5<sup>th</sup>,  $n=118$ , 150 and 101, respectively) the values predicted by the logistic model were bigger than the observed values. The HL test was statistically significant ( $<0.0001$ ). However, both alternative measures indicated good fit of the model.

#### *Discrimination*

In terms of discriminative ability, the modified G8 showed a good performance, with a c-statistic value of 0.85 (95% CI: 0.82–0.87). Sensitivity and specificity were respectively 82.2% and 69.1%.

**Table 4.4.** Performance of the modified G8 in development and validation datasets

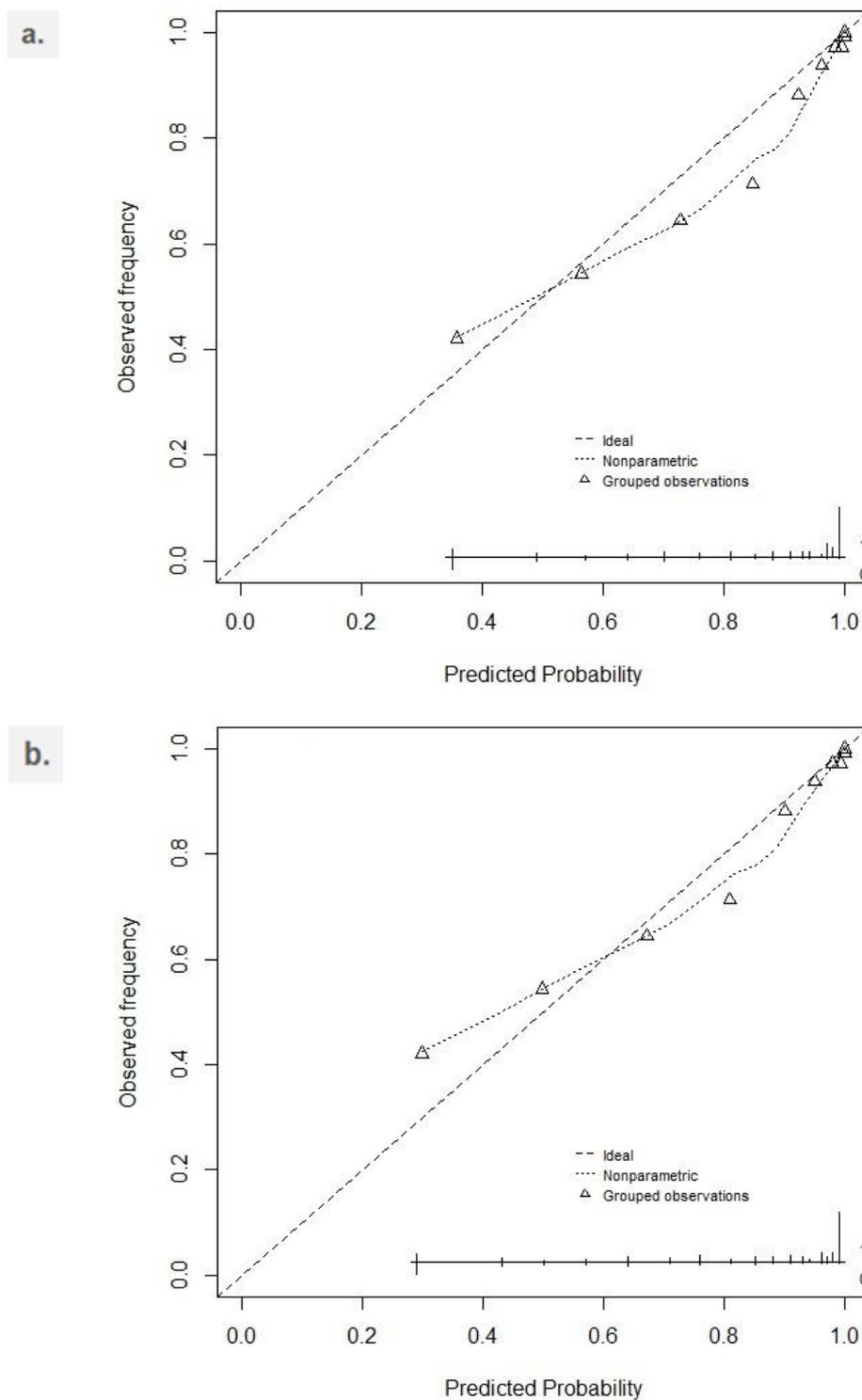
Model performance	Development set	Validation set	
		Original model	Updated model <sup>b</sup>
<b>Overall performance</b>			
Brier score	0.07	0.12	0.12
<b>Discrimination</b>			
C-statistic (95% CI)	0.92 (0.89-0.94)	0.85 (0.82-0.87)	Same discrimination as the original model
Se (95% CI)	89.2% (86.5-91.5)	82.2% (79.8-84.5)	
Sp (95% CI)	79.0% (69.4-86.6)	69.1% (63.3-74.6)	
PPV (95% CI)	96.5% (94.7-97.9)	91.1% (89.1-92.9)	
NPV (95% CI)	52.8% (44.3-61.2)	50.3% (45.1-55.5)	
LR+ (95% CI)	4.24 (2.87-6.26)	2.66 (2.22-3.19)	
LR- (95% CI)	0.14 (0.11-0.18)	0.26 (0.73-0.79)	
<b>Calibration</b>			
HL P-value	0.97	<b>&lt;0.0001</b>	<b>0.0002</b>
Pearson $\chi^2$ P-value	0.29	0.985	0.984
Stukel test P-value	0.11	0.551	0.496
Slope (95% CI)	1 * (0.79;0.21)	0.74 (0.64;0.85)	0.74 (0.64;0.85)
Intercept (95% CI) <sup>a</sup>	0 * (-0.27;0.27)	-0.25 (-0.41;-0.08)	0 (-0.17;0.17)
MAPE	0.03	0.08	0.06

Abbreviations: CI, confidence intervals; Se, sensitivity; Sp, specificity; PPV, positive predictive value; NPV, negative predictive value; HL, Hosmer-Lemeshow test; MAPE, mean absolute prediction error.

<sup>a</sup> Calibration intercept with calibration slope fixed at 1.

<sup>b</sup> Recalibration of the intercept:  $-0.247 + \text{Linear Predictor}$  from the original model

\* Perfect values by definition.



**Figure 4.1.** Calibration plots of the modified G8 applied in the independent cohort (ONCODAGE)

Predicted versus observed probabilities of the a) original model; b) updated model (recalibration of the intercept). Triangles indicate the observed frequencies by decile of predicted probabilities. The distribution of subjects is indicated with spikes at the bottom of the graph, stratified by the outcome (abnormal GA: 1 / normal GA: 0).

## 4.3.4 Interpretation of model validation results and updating

Both samples were poorly related in terms of the considered predictors and outcome (abnormal GA), and main baseline patient characteristics, as shown in **Table 4.2**. As a consequence of very different rates of the outcome prevalence between the development and the validation cohorts, the difference between observed and predicted outcome frequency in the validation set were significantly different (79% vs. 82%, respectively), which deteriorates the calibration in the large. The increased variability of the LP further indicated the heterogeneity of case mix, which can explain the different discriminative ability of the modified G8 in the validation cohort. These results led us to update the model, adjusting the mean predicted probability so that it equals the observed outcome frequency.

The model with an adjusted intercept was also tested in the validation set (intercept of -0.74 instead of -0.49). **Figure 4.1(b)** presents the calibration plot of the updated model, showing better calibration. **Table 4.5** presents additionally the observed versus predicted probabilities by decile of risk of the original and updated model, showing for the latter only small deviations from perfect fit. Note that when risks were low, the predicted probabilities were slightly too low (1<sup>st</sup> decile of risk).

**Table 4.5.** Observed vs. Predicted probabilities of an abnormal GA, by decile of risk

Group	Observed mean	Original model		Updated model	
		Mean PP	Difference <sup>a</sup>	Mean PP	Difference
1	0.42	0.36	-0.06	0.31	-0.11
2	0.54	0.56	0.02	0.50	-0.04
3	0.64	0.72	0.08	0.67	0.03
4	0.71	0.84	0.13	0.81	0.10
5	0.88	0.92	0.04	0.90	0.02
6	0.94	0.96	0.02	0.95	0.01
7	0.97	0.98	0.01	0.98	0.01
8	0.97	0.99	0.02	0.99	0.02
9	0.99	1	0.01	1	0.01
10	1	1	0	1	0
Total population	0.79	0.82	0.03	0.79	0

PP, predicted probability.

<sup>a</sup> Difference between Observed vs. Predicted probabilities

Overall, most calibration measures improved with the updated model. For example, even if the MAPE of both, the original and updated model, were inferior in the validation cohort compared to that of the development cohort, the error of the updated model was smaller (0.06 vs. 0.08).

#### 4.3.5 AUROC comparisons

- Modified G8 in de validation set vs. development set, by cancer site

First, when we compared the performance of the modified G8 in each cohort by cancer site, significant differences were found for prostate and breast cancer, with higher AUROC in the development cohort (**Table 4.6**).

- Original versus Modified G8, overall and by cancer site

Overall, the AUROC of the modified G8 was higher compared to the original tool ( $p=0.019$ ), with values of 84.6% (82.3-86.8) and 81.7% (95% CI: 79.3-84.1), respectively (**Figure 4.2**).

When both screening tools (original and modified G8) were compared in the external validation set, by the 5 main cancer sites, they yielded not significantly different results, except for the AUROC of the modified G8 in hematological malignancies, which was significantly higher than the AUROC of the original G8 ( $p=0.019$ ) (**Table 4.6**).

**Table 4.6.** Comparisons of AUROC in main tumor sites: modified G8 in de validation set vs. development set; original vs. modified G8 in the validation set

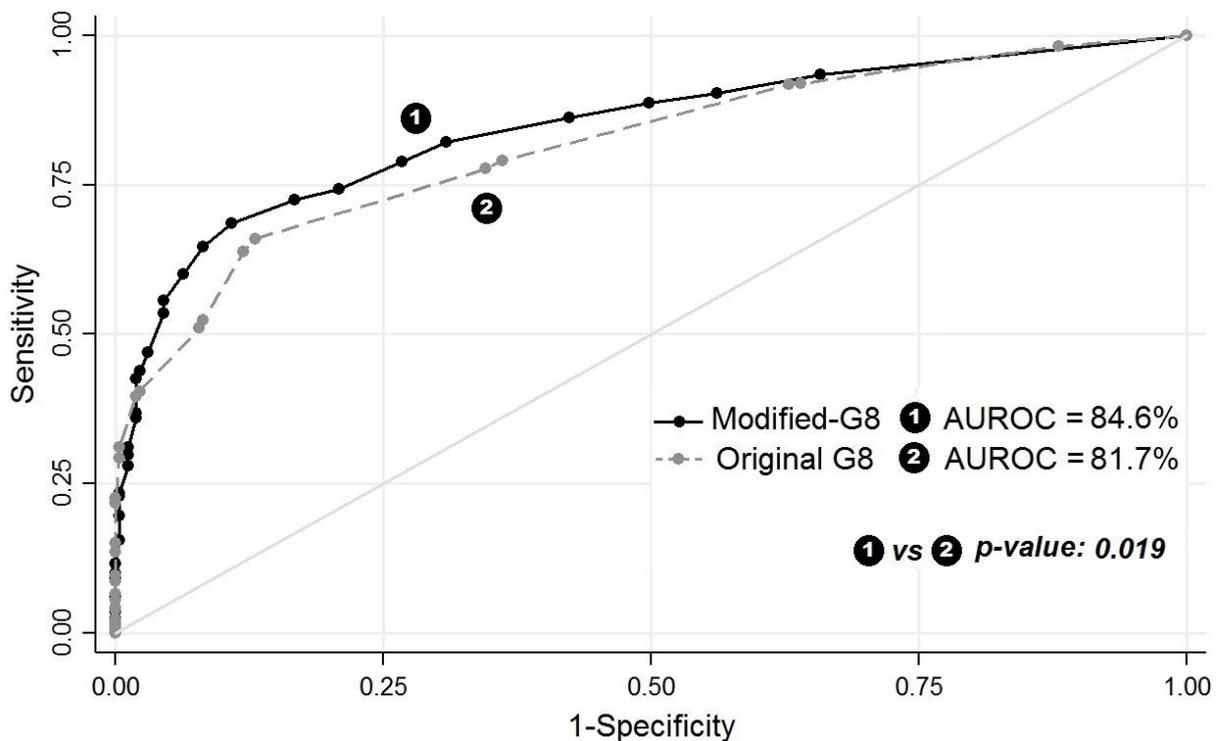
Tumor site	Modified G8		P-value <sup>a</sup>	Original G8	
	Development set	Validation set		Validation set	P-value <sup>b</sup>
Colorectal	0.87 ± 0.04	0.85 ± 0.04	0.755	0.80 ± 0.05	0.252
Prostate	0.87 ± 0.03	0.75 ± 0.04	<b>0.026</b>	0.76 ± 0.05	0.829
Breast	0.91 ± 0.02	0.82 ± 0.02	<b>0.001</b>	0.79 ± 0.02	0.091
Lung	*	0.78 ± 0.07	-	0.84 ± 0.04	0.154
Hematologic	0.91 ± 0.04	0.89 ± 0.04	0.797	0.80 ± 0.04	<b>0.019</b>

sd, standard deviation.

\* All patients had an abnormal GA (n=17)

<sup>a</sup> AUROC of the modified G8 in the validation set versus AUROC in the development set.

<sup>b</sup> AUROC of the original G8 versus modified G8 in the external validation set.



**Figure 4.2.** ROC curves for predicting impairment of the geriatric assessment, used as the reference test: original versus modified G8 questionnaire (external validation set; n = 1304)

Abbreviations: AUROC, area under the ROC curve; ROC, receiver-operating characteristic.

## 4.4 Discussion

### 4.4.1 Summary of findings

We externally validated the modified G8 in a large multicenter prospective cohort of older patients with cancer. Since the main objective of this thesis was to develop an optimized version of the G8 to improve the identification of older patients with cancer requiring a full GA, it was crucial to confirm its diagnostic performance in an independent cohort consistent with the target population.

The modified G8 demonstrated very good discriminative ability in the validation cohort, with a c-statistic of 0.85 [95% CI: 0.82–0.87]. Yet and as expected in external validation studies, this performance was significantly lower compared to that of the development cohort (0.92 [0.89-0.94]). Different points should be discussed with this respect. First, because both samples had substantially different distribution of case mix, different model performances were to be expected [130]. Second, one particular individual regression coefficient may have been incorrectly estimated in the validation cohort due to the impossibility of characterizing this predictor as in the development study. Because the item corresponding to past history of heart failure or coronary heart disease was unavailable in the ONCODAGE cohort, the category “heart” of the CIRS-G was used as a proxy, which includes ischemic disorders and heart failure, but also several other and unrelated heart problems such as complete arrhythmia with atrial fibrillation or other heart rhythm disorders. Importantly, those cardiac disorders were already shown to have inferior or no association with an abnormal GA in the univariate analysis performed in the development of the modified G8 (i.e. heart rhythm disorder: OR=0.69 [0.41-1.17]; p=0.167). To further assess the impact of this substitution, we

retrospectively entered the CIRS-G 'Heart' category in the multivariate model from the development study, and found a similar and not statistically significant coefficient to that of the validation cohort (0.11 [-0.01-0.24];  $p=0.074$ ). In the same way, a decreased specificity was shown when this category was used as part of the modified G8 in the development cohort, from 79% to 69%, the same specificity observed in the validation cohort. All these elements suggest a probably underestimated performance of the modified G8 in the present study and that improvements in accuracy could likely be achieved when using correct scoring.

In terms of calibration, we first found a significant value for the Hosmer-Lemeshow test. This test is frequently used to assess calibration in prediction models, however, several deficiencies have been pointed out [136-140]: 1) it does not provide a measure of the magnitude of miscalibration; 2) the test is highly sensitive to sample size: with large sample sizes, any slight deviations in calibration will yield a statistically significant result; and 3) results can differ markedly depending on the arbitrary choice of number of risk groups. Therefore, two alternatives were considered: a standardized Pearson chi-square to evaluate the discrepancy between predicted and observed outcomes, and the Stukel test, that is not a calibration test in the sense of explicitly comparing observed to predicted outcomes based on the model, but instead creates two new parameters based on the linear predictor from the fitted model and added them to test the null hypothesis that both of their coefficients are equal to 0. The latter has been referred as a very powerful test [139,141]. Even when the HL test indicated poor calibration, these two other measures indicated well fitted of the model in the validation sample. Second, the intercept of -0.27 indicated that predictions were on average too high, which was related to a lower percentage of identified cases of an abnormal GA in the current study (79.4%), compared to the development study (86.7%). This difference could lead to a

reduced calibration. When an adjustment of the intercept was performed, calibration of the model improved as shown in the calibration plot, where the predicted probabilities were on average similar to the observed frequencies. This showed that a simple recalibration method could improve the performance of the instrument when applied in new patients.

We further compared the diagnostic performance of this modified G8 tool to the original G8, which was externally validated previously in this same population [85]. Overall, the modified G8 showed better discriminative ability with an AUROC of 84.4% compared to 81.7% (original G8). According to the most recent review of screening tools warranting a GA in older patients with cancer [56], the most studied instruments were the fTRST (2 studies), the G8 (8 studies) and the VES-13 (11 studies). The 2 first instruments had the highest sensitivities, while the VES-13 had the highest specificity. Median sensitivities for the G8 and the VES-13 were respectively 85% and 62%, while median specificities were 59% and 78%. As for the fTRST, when using the cutoff of  $\geq 1$  for the older cancer population, it led to sensitivities of 92% and 91% and specificities of 42% and 50%. The modified G8 had a very good performance with a sensitivity of 82% and a specificity of 69%. By cancer site, the original and modified G8 had similar discriminative performances, with the exception of patients with hematological malignancies, where the modified G8 had better discrimination.

#### 4.4.2 Strengths and limitations

This study has several strengths that should be noted. First, the same population used to externally validate the original G8, was used to validate the modified G8, where their performances were assessed and directly compared. Second, we followed a

standardized and documented framework [130,142] for external validation studies that consisted briefly in the following steps : 1) comparing the development and validation populations, 2) evaluating model performance by main measures of discrimination *and* calibration, and 3) interpreting the results with eventually updating of the model. The two latter points are of particular importance compared to other validation studies. According to a systematic review concerning methods used in external validation studies of multivariable prediction models [124], calibration was rarely assessed (35%), even when it is an important and widely recommended measure [110,131,142-145]. Moreover, the majority of the studies reported only the Hosmer-Lemeshow test as a measure of calibration, with no graphical representation of predicted and observed probabilities. In the TRIPOD document (Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis) it was also stated that very few studies developing or validating prediction models for the same outcome compare performance against other existing models [142]. Decoster et al. [56], found only 11 studies that directly compared two or more screening tools with GA, and reported only on the discriminative ability, with no report on calibration measures. Furthermore, in our study, we reported results overall and by tumor site comparing the original and modified G8, and comparing the modified G8 in the development and validation cohorts.

Some limitations should also be mentioned. Among the 7 scales of the GA (reference standard) used in the development cohort, one was not available in the validation dataset, for which a more accurate scale was used as a substitute. Still, they both measure the same domain of the GA evaluating mood; the 4-item version was used in the development cohort and the 15-item version, in the external validation cohort. Similarly and as previously discussed, one item of the modified G8 was not available in the validation dataset and thus substituted by an imperfect proxy, which could have led

to misclassification bias. Finally, 131 patients were excluded from the analysis as they had at least one missing value on either the items of the modified G8 or the GA. However, main patient's characteristics and outcome were similar compared to the included patients (i.e. age, cancer site and metastatic status), with only small differences concerning gender and ECOG-PS.

## **4.5 Conclusion**

In this external validation study, the modified G8 confirmed its good discriminative ability, while calibration was improved by simple recalibration of the model to the prevalence of the target population. These results support the utility and generalizability of the instrument to other related populations.

## CHAPTER 5

---

### **Varying Gold Standard definitions**

## **5.1 Introduction**

The term “gold standard” or “reference standard” is used to describe the best available method for establishing the presence or absence of a condition or disease of interest [146], and thus constitutes the ultimate measure for comparison of new diagnostic or screening tests in test accuracy studies. Yet, this assumes that an established gold standard is available and has perfect accuracy, but this is not always the case. Gold standard tests for many diseases may be difficult to implement due to their invasiveness or may lack 100% accuracy or a clear cut-off value on the reference standard [147,148]. In other cases, there is no unequivocal definition available for the disease or target condition, preventing the characterization of a clear and definite gold standard. No consideration of such gold standard imperfections may lead to erroneous conclusions regarding the accuracy and clinical utility of novel tests under study.

In order to detect health problems in older patients with cancer and tailoring treatment decisions accordingly, a multidimensional GA is recommended [27]. Because GA is a time-consuming process and requires specific expertise for its conduction, screening tools have been developed to help identifying potentially frail patients warranting a complete GA. However, there is no unique definition of what constitutes this population and what the reference gold standard should be. There is currently no consensus for defining and measuring frailty and yet no broadly accepted standard for classification of older cancer patients according to their health status. Several classifications usually based on clinical expertise and professional consensus have been used, but their concordance was found variable with different patients being identified as frail depending on the criteria used [149].

In the geriatric oncology setting, a pragmatic definition based on  $\geq 1$  abnormal test at the GA has been so far mostly used for developing and validating screening instruments [56,61,122], but this approach is hampered by a lack of standardization in GA components across studies. This definition also does not capture important aspects of the reality of clinical practice in geriatric oncology, such as actual treatment decisions based on GA findings, expert-based clinical classifications and/or broader approaches to frailty.

In this context, we aimed to measure the impact of varying gold standard definitions on the diagnostic performance of two screening tools specifically developed for older patients with cancer, the G8 [61] and the modified G8 [122]. To do so, we assessed the predictive performance of the G8 instruments under six different classifications and definitions evocative of a state of frailty.

## 5.2 Methods

### 5.2.1 Study population

The ELCAPA prospective cohort is described in Chapter 2 (General methodology).

**Figure 2.3** presents the flow chart of participants included in this study.

### 5.2.2 Reference standard definitions

The following reference standard definitions evocative of a state of frailty were tested: a) detection of  $\geq 1$  or b) 2 impaired components of the GA; c) prescription of  $\geq 1$  clinically significant intervention by the geriatrician; d) identification of a vulnerable profile as defined by a latent class approach [150] or e) by expert-based classifications from Balducci [48] and f) Droz [49] (entitled SIOG classification in our study).

### ***Geriatric assessment***

The GA included a variety of domains covering functional status, mobility, nutrition, cognition, mood and comorbidities, used in the development of the modified G8 screening tool [122] and in accordance with international recommendations [28]. Domains were evaluated by the following validated tests: Activities of Daily Living (ADL $\leq$ 5/6), Instrumental Activities of Daily Living (IADL $\leq$ 7/8), Mini Mental State Examination (MMSE $\leq$ 23/30), mini-Geriatric Depression Scale (mini-GDS $\geq$ 1), Mini Nutritional Assessment (MNA $\leq$ 23.5/30), Cumulative Illness Rating Scale for Geriatrics (CIRS-G; at least one comorbidity grade 3 or 4), and Timed Up-and-Go test (TUG $>$ 20s). Thresholds considered:  $\geq$ 1 and  $\geq$ 2 impaired components.

### ***Geriatric interventions***

For each patient, proposed geriatric interventions after GA were documented. After internal review by two expert geriatricians (ML, PC) and for the sake of the present analysis, five domains covering clinically relevant deficiencies that may warrant further geriatric interventions were distinguished: nutritional support (including dietary advice and nutritional supplements), home care (including nursing and physiotherapy), neuropsychological support, social support and adaptation of the anticancer treatment. A consideration of  $\geq$ 1 of these interventions prescribed by the geriatrician was defined as reference standard.

### ***Frailty classifications***

Three classifications were considered to approach the non-standardized definition of frailty: Balducci's, SIOG and a latent class typology (LCT), using the "unhealthy" profiles as reference standards.

**Table 5.1** details the indicators considered to categorize patients as fit or “unfit” (vulnerable or frail or too sick) according to Balducci and SIOG classifications.

Following Balducci et al. [48] classification, fit patients were defined as those functionally independent (no dependence in ADL) and without severe comorbidity (CIRS-G grade 0, 1 or 2) and without geriatric syndromes, and unfit patients, as those aged over 85 years and/or dependent in one or more ADL ( $\leq 5/6$ ) and/or with one or more severe comorbid conditions (CIRS-G grade 3 or 4) and one or more geriatric syndromes. Dementia (MMSE $\leq 23/30$ ), delirium, depression (from the Diagnostic and Statistical Manual of Mental Disorders [DSM-IV]), urinary and/or fecal incontinence and falls were considered geriatric syndromes. Some geriatric syndromes considered in the original definition for frailty according to Balducci were not available in the database and were not taken into account (i.e. osteoporosis, neglect and abuse, and failure to thrive). The qualification for falls according to Balducci is  $\geq 3$  times per month, although it was different in our database, define as  $\geq 1$  falls in the last 6-months.

Regarding the SIOG classification [49], patients with no serious comorbidity (CIRS-G grade 0, 1 or 2), functionally independent (no dependence in IADL and ADL), and without malnutrition were considered as fit, whereas patients with dependency in one or more ADL ( $\leq 5/6$ ) or IADL ( $\leq 7/8$ ) and/or one or more severe comorbid conditions (CISR-G Grade 3-4) and/or malnutrition were considered as unfit. The original definition for malnutrition was not available in our database, so we used the following substitute of the variable, according to French guidelines [151]:  $\geq 5\%$  of weight loss in the last month and/or  $\geq 10\%$  within the last 6 months instead of  $\geq 5\%$  during the previous 3 months.

**Table 5.1** Frailty classifications approaches (fit vs. "unfit" patients) and indicators.

Classifications	Indicators	Fit	Unfit <sup>a</sup>
<i>Balducci et al., 2000</i>	Age	≤ 85 yrs and	> 85 yrs and/or
	ADL (Katz)	6/6 and	≤5/6 and/or
	Comorbidities grade 3-4	0 and	≥1 and/or
	Geriatric syndromes <sup>b</sup>	0	≥ 1
<i>Droz et al. (SIOG), 2010</i>	ADL (Katz)	6/6 and	≤ 5/6 and/or
	IADL (Lawton) <sup>c</sup>	8/8 and	≤ 7/8 and/or
	Comorbidities grade 3/4	0 and	≥1 and/or
	Malnutrition <sup>d</sup>	absence	at risk or severe

<sup>a</sup> vulnerable or frail or too sick

<sup>b</sup> among dementia (MMSE≤23/30), delirium, depression (diagnosed by a semi-structured interview to identify criteria for a major depressive episode from the Diagnostic and Statistical Manual of Mental Disorders [DSM-IV]), urinary and/or fecal incontinence and falls (≥ 1 fall in the last 6-months).

<sup>c</sup> in sensitivity analysis we considered the 4-item IADL for men: ability to manage money, to manage medications, to use transportation, and to use the telephone.

<sup>d</sup> absence: <10% of weight loss in the past 6-months and <5% in the last month; at risk : 10-15% of weight loss in the past 6-months and/or 5-10% in the last month ; severe malnutrition: ≥15% of weight loss in the past 6-months and/or ≥10% in the last month.

Additionally to these two classifications, we considered a LCT recently developed in a population of older patients with cancer, combining components of the GA [150]. We applied the scoring rules to classify patients into one of the four profiles identified (relatively healthy, malnourished, cognitive and mood impaired, and globally impaired). The scoring equations were based on a set of indicators and covariates yielding posterior class membership probabilities for each patient. A patient was categorized as “fit” if its membership probability to Class 1 was ≥50% and “unfit” if that probability was <50%. Variables used in the algorithm to classify patients are reported in **Table 5.2**.

### 5.2.3 Screening tools

The original G8 [61] is presented in **Table 1.6** (Chapter 1) and the modified G8 [122], in **Table 3.4** (Chapter 3).

**Table 5.2** Variables and definitions used to classify patients in the latent class typology

Variable	Definition
Inadequate social environment	Absence of a primary caregiver or of adequate support at home or of a strong circle of family and friends able to meet the needs of the patient at the time of the evaluation
Malnutrition	One or more of the following criteria as recommended by the French National Authority for Health: at least 10% weight loss in 6 months or 5% in 1 month and/or body mass index less than 21 kg/m <sup>2</sup> and/or Mini-Nutritional Assessment (MNA) score less than 17/30 and/or serum albumin level less than 35 g/L [151]
Depression	Diagnosed by a semi-structured interview to identify criteria for a major depressive episode from the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) [103]
Cognitive impairment	Mini-Mental State Examination score (MMSE) $\leq 23/30$ [101]
Number of severe comorbidities (grade 3-4)	As assessed by the Cumulative Illness Rating Scale for Geriatrics (CIRS-G; 0,1, $\geq 2$ ) [106]
Functional impairment	Activities of Daily Living score (ADL) $\leq 5/6$ [97]
Age	In two classes: $\leq 80$ years; $> 80$ years
Tumor site	Colorectal, upper gastrointestinal tract and liver, breast, prostate, other urologic malignancies, hematologic malignancies, other
Metastatic status	M0, absence of distant metastases; M1, presence of distant metastases; Mx, metastatic status unknown
Status at the time of the GA	In/outpatient

#### 5.2.4 Statistical analysis

The study population was described in terms of clinical, demographic characteristics and geriatric assessment results. Univariate logistic regression analyses were used to assess the associations between the different reference standards and both screening tools. A test for equality of the regression coefficients of both tools was performed. Areas under the Receiver Operating Characteristic (AUROC) curves were calculated to compare the diagnostic performance of both screening tools against

different gold standards. A test for the equality of the AUROC using an algorithm suggested by DeLong and Clarke-Pearson[152] was carried out for comparison of both curves. Their 95% confidence intervals (95% CI) are also reported. We additionally investigated whether a different cut-off value provided a better discriminative performance for each reference standard. Sensitivities and specificities were calculated for optimal cut-off values (those prioritizing sensitivity), along with their 95% CI and were compared using McNemar's Chi-square test. In additional analysis, chi-square tests were used to compare variables between the two groups built from the LCT.

### **Sensitivity analyses**

In order to have a homogenous population and compare both screening tools against the different reference definitions, patients with missing data on any of the definitions or G8 scores were excluded from the analysis. To test the robustness of our results we performed each comparison including patients with available G8 scores and the definition tested, when other definitions were missing.

The SIOG classification, considered for older patients with prostate cancer, uses a 4 item IADL: ability to manage money, to manage medications, to use transportation, and to use the telephone. Thus, we additionally considered the following categorization in sensitivity analyses: abnormal IADL  $\leq 7/8$  for women and  $\leq 3/4$  for men.

## 5.3 Results

### 5.3.1 Study population

Between January 2007 and June 2015, 1943 patients were included in the ELCAPA cohort, of whom 1136 had complete data for the G8 and modified G8 and available information for each of the six reference definitions tested. Compared to the population not included in the analysis (n=807), included patients had more pejorative scores G8 and modified G8 and a higher proportion of patients had an impaired MNA (66% vs. 58%, p=0.001). In contrast, there were more patients with abnormal tests ADL and IADL in the non-included population. More details in Annex 3, **Table A3.3**.

#### *Patient characteristics and geriatric interventions*

Main patient characteristics and results from GA of the study population are shown in **Table 5.3**. Median age was 80 years (IQR: 76-85). The most frequent cancers were those of the digestive system (36.3%), followed by breast cancer (16%), and urinary tract cancer (14.8%), with almost half of the patients presenting metastasis (43.5%). A loss of functional capacities was common, with respectively 31.6% and 58.5% of patients having at least one impairment for ADL and IADL scales. Malnutrition was identified in 68.3% of patients according to the French National Health Authority criteria, whereas an impaired MNA was identified in 64.3% of patients, indicating a prevalent nutritional impairment in the population study. The burden of comorbidities was high with 63.1% of patients having at least one comorbidity of severity grade 3 or 4 according to CIRS-G criteria.

**Table 5.3.** Patient characteristics

Characteristics (N=1136)	N	%
Age in years, median (IQR)	80 (76;85)	
Number of medications/d, median (IQR)	6 (4 ; 8)	
Outpatient	412	36.3
Male gender	587	51.7
Cancer type		
Colorectal	201	17.7
Liver or upper gastrointestinal tract	211	18.6
Urinary tract	168	14.8
Prostate	127	11.2
Hematological	84	7.4
Breast	182	16.0
Others <sup>a</sup>	163	14.3
Metastasis	494	43.5
Inappropriate Social environment <sup>b</sup>	177	15.6
Functional impairment		
ADL ≤5	359	31.6
IADL ≤7	665	58.5
ECOG Performance Status		
0: Fully active	205	18.0
1: Restricted activity	342	30.1
≥2: Unable to carry out work activities/confined to bed>50% or disabled	589	51.9
Cognitive impairment (MMSE ≤23)	285	25.1
Depressive disorder		
Mini-GDS ≥1	379	33.4
DSM IV criteria	364	32.0
Malnutrition		
MNA ≤23.5	730	64.3
HAS criteria <sup>c</sup>	776	68.3
At risk or severe malnutrition <sup>d</sup>	317	27.9
Comorbidities		
CIRS-G (≥1, grade 3/4)	717	63.1
≥1grade 3 comorbidities	675	59.4
≥1grade 4 comorbidities	182	16.0
Mobility		
TUG ≥20s	434	38.2
Fall during the previous 6 months	365	32.1

Abbreviations: IQR, interquartile range; M0, absence of distant metastases; M1, presence of distant metastases; Mx, metastatic status not assessable; NA, not applicable; ECOG-PS, Eastern Cooperative Oncology Group-Performance Status; ADL, Activities of Daily Living; IADL, Instrumental ADL; MMSE, Mini-Mental State Evaluation; GDS, Geriatric Depression Scale; DSM, Diagnostic and Statistical Manual of Mental Disorders; MNA, Mini Nutritional Assessment; HAS, French National Authority for Health; CIRS-G, Cumulative Illness Rating Scale for Geriatrics; TUG, Time get up and go.

<sup>a</sup> Lung (n=44), skin (n=32), unknown primary origin (n=30), sarcoma (n=15), gynecologic (n=14), brain (n=11), head and neck (n=5), thyroid (n=3), others (n=9)

<sup>b</sup> Defined as absence of a primary caregiver or of adequate support at home or of a strong circle of family and friends able to meet the needs of the patient at the time of the evaluation.

<sup>c</sup> One or more of the following criteria: at least 10% weight loss in 6 months or 5% in 1 month and/or body mass index <21 kg/m<sup>2</sup> and/or Mini-Nutritional Assessment score <17/30 and/or serum albumin <35 g/L.

<sup>d</sup> Weight loss ≥10% in the last 6 months and/or ≥5% in the last month

**Table 5.4** details the interventions prescribed by the geriatrician after GA. A median of 3 interventions (IQR 2-5) were proposed for each patient. The most frequent one concerned nutritional support (74.5%), while physiotherapy and social support were proposed for 63.8% and 63.5% of patients, respectively. More infrequently prescribed interventions were nursing (6.8%) and psychiatric care (4.1%). Overall, at least one intervention was proposed by the geriatrician in most patients (1032; 90.9%).

**Table 5.4.** Geriatric interventions for overall patient management

<b>Geriatric interventions</b>	<b>N</b>	<b>%</b>
<b>≥ 1 Nutritional support</b>	846	74.5
Dietary advice	744	65.5
Nutritional supplements	353	31.1
<b>≥ 1 Home care</b>	741	65.2
Physiotherapy	725	63.8
Nursing	77	6.8
<b>≥ 1 Social support</b>	721	63.5
Social care	665	58.5
Personal assistance	315	27.7
Personal care allowance (APA)	224	19.7
<b>≥ 1 Neuropsychological support</b>	438	38.6
Psychological care	431	37.9
Psychiatric care	46	4.1
<b>Adaptation of the anticancer treatment</b>	263	23.2
<b>≥ 1 intervention prescribed</b>	1032	90.9

### 5.3.2 Prevalence of frailty by reference standard

The percentage of patients classified as frail according to the different reference standards varied as follows: 76.9% (GA: ≥2 impairments), 79.5% (LCT), 83.2% (SIOG classification), 86.5% (Balducci's classification), 91.9% (GA: ≥1 impairment).

With respect to the LCT classification, comparison between fit and unfit profiles revealed the latter as more likely to have functional and cognitive impairment, malnutrition, an inadequate social environment, depression, and more severe comorbidities than the fit group (p-values <0.05; Annex 7, **Table A7.1**).

### 5.3.3 Predictive performance of screening tools by reference standard

In univariate logistic regression analyses (**Table 5.5**), abnormal G8 screening scores were significantly associated with all reference standard regardless of the definition used. Nutritional support had the strongest association among the types of interventions (Original G8: OR 8.6 [95% CI 5.9-12.6]; Modified G8: OR 9.1 [6.3-13]). Similar OR's were found between both tools, however, significant differences in favor of the modified G8 were found concerning GA  $\geq 1$  impairment,  $\geq 1$  geriatric intervention prescribed and SIOG classification (*chi-square p-values*: 0.0026, 0.0219 and 0.0069 respectively).

**Figure 5.1** shows the comparisons of the ROC curves of both scores for each reference standard. AUROC were equal or higher than 0.80 for both tools and all definitions tested. Comparing the two instruments, AUROC were significantly higher in favor of the modified G8 to predict 4 out of the 6 definitions tested: at least one impairment in GA (modified G8: 0.93 [95%CI 0.91–0.95] vs. original G8: 0.90 [0.87–0.92]; p=0.0029), two or more impairments in GA (modified G8: 0.90, [0.88–0.92] vs. original G8: 0.87 [0.88–0.92]; p=0.0006), at least one geriatric intervention prescribed (modified G8: 0.85 [0.81–0.89] vs. original G8: 0.81 [0.77–0.86]; p=0.0056) and unfit patients according to SIOG classification (modified G8: 0.88 [0.86–0.91] vs. original G8: 0.83 [0.81–0.86]; p<0.00001). No significant difference was found for the LCT and Balducci's classification.

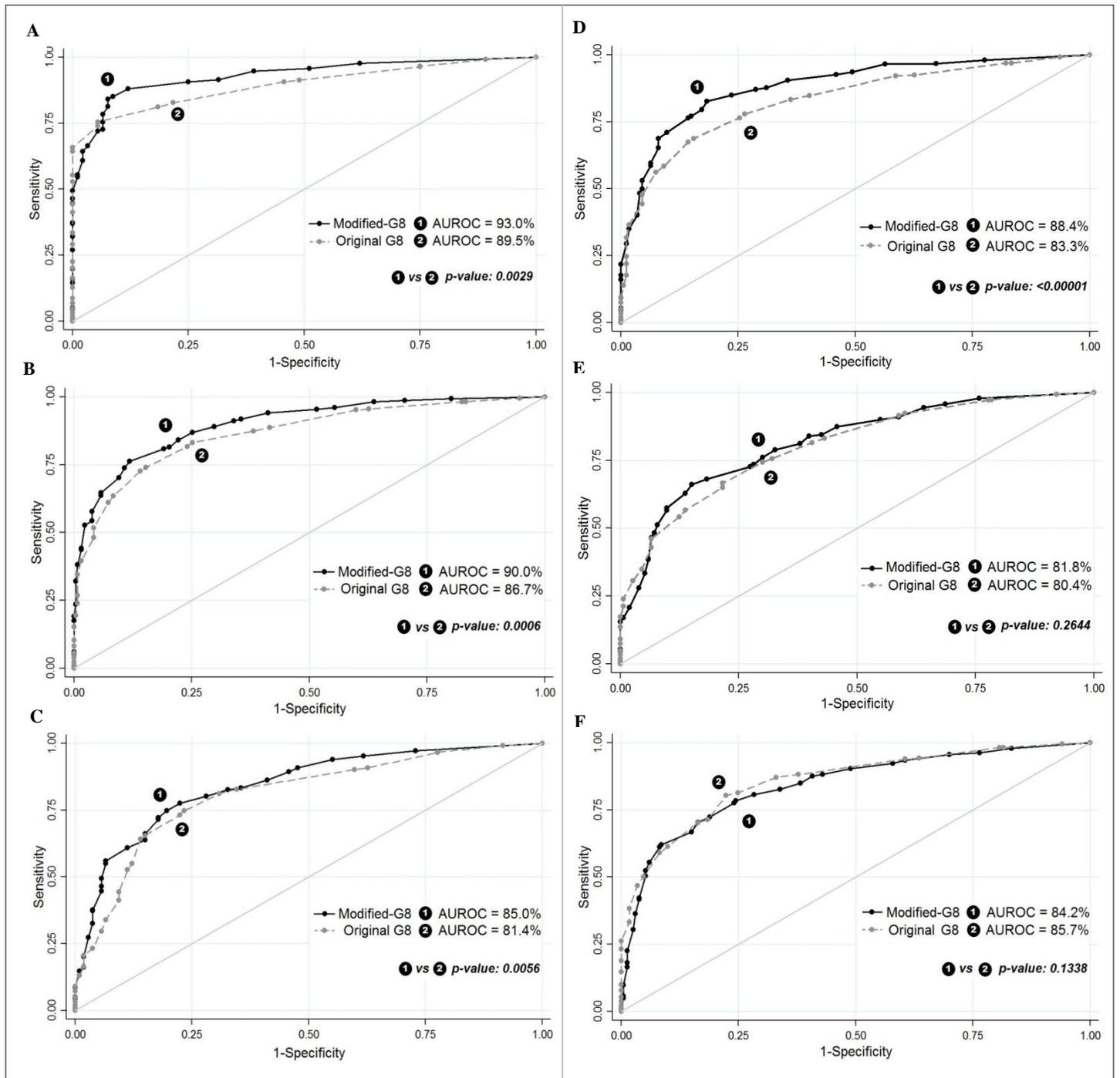
**Table 5.6** details the diagnostic performances of each tool for the six definitions tested. Sensitivities based on optimal cutoffs were of similar magnitude for both tools, ranging from 83% (original G8) and 85% (modified G8) for Balducci's classification, to 91% (both tools) for  $GA \geq 1$  impairment, although significant differences were found for  $GA \geq 2$  impairments and SIOG classification in favor of the modified G8. Most specificities were higher for the modified G8. They ranged from 41% ( $\geq 1$  intervention prescribed) to 62% (Latent class typology) for the original G8, and from 56% ( $\geq 1$  intervention prescribed) to 75% ( $GA \geq 1$  impairment) for the modified G8.

**Table 5.5.** Associations between different proposals of reference standard and both screening tools (original G8 and modified G8), N=1136

Reference standards	Screening tools	Normal Scores <sup>a</sup>	Abnormal Scores <sup>b</sup>	OR (95% CI)	p-values
		N (%)	N (%)		
<b>GA ≥ 1 impairment</b> (Abnormal: n=1044)	Original G8	97 (9.3)	947 (90.7)	11.62 (7.33-18.42)	<0.0001
	Modified G8	98 (6.4)	946 (90.6)	28.96 (17.29-48.50)	<0.0001
<b>GA ≥ 2 impairments</b> (Abnormal: n=874)	Original G8	42 (4.8)	832 (95.2)	13.25 (8.91-19.69)	<0.0001
	Modified G8	40 (4.6)	834 (95.4)	19.61 (13.16-29.24)	<0.0001
<b>Geriatric Interventions</b> (≥ 1: n=1032)	Original G8	104 (10.1)	928 (89.9)	6.29 (4.05 - 9.76)	<0.0001
	Modified G8	109 (10.6)	923 (89.4)	10.68 (6.91-16.49)	<0.0001
<b>≥1 Nutritional support</b>	Original G8	48 (5.7)	798 (94.3)	8.62 (5.90-12.59)	<0.0001
	Modified G8	55 (6.5)	791 (93.5)	9.05 (6.30-12.99)	<0.0001
<b>≥1 Home care</b>	Original G8	65 (8.8)	676 (91.2)	2.72 (1.92 - 3.87)	<0.0001
	Modified G8	53 (7.2)	688 (92.8)	5.27 (3.70 - 7.50)	<0.0001
<b>≥1 Social support</b>	Original G8	64 (8.9)	657 (91.1)	2.57 (1.81 - 3.65)	<0.0001
	Modified G8	74 (10.3)	647 (89.7)	2.53 (1.81 - 3.52)	<0.0001
<b>≥1 Neuropsychological support</b>	Original G8	26 (5.9)	412 (94.1)	3.32 (2.14 - 5.17)	<0.0001
	Modified G8	33 (7.5)	405 (92.5)	2.92 (1.95 - 4.36)	<0.0001
<b>Treatment adaptation</b>	Original G8	27 (10.3)	236 (89.7)	1.39 (0.89 - 2.17)	0.142
	Modified G8	25 (9.5)	238 (90.5)	1.84 (1.18 - 2.90)	0.007
<b>SIOG classification</b> (Unfit <sup>c</sup> : n=945)	Original G8	74 (7.8)	871 (92.2)	8.31 (5.66-12.20)	<0.0001
	Modified G8	69 (7.3)	876 (92.7)	14.92 (10.14-21.94)	<0.0001
<b>Balducci's classification</b> (Unfit <sup>d</sup> : n=983)	Original G8	84 (8.5)	899 (91.5)	7.49 (5.06-11.09)	<0.0001
	Modified G8	98 (10)	885 (90)	7.42 (5.07-10.85)	<0.0001
<b>Latent class typology</b> (Unfit <sup>e</sup> : n=903)	Original G8	55 (6.1)	848 (93.9)	10.06 (6.89-14.69)	<0.0001
	Modified G8	69 (7.6)	834 (92.4)	8.77 (6.14-12.55)	<0.0001

GA, Geriatric Assessment; OR, Odds Ratio; 95% CI, 95% confidence interval

<sup>a</sup> G8: >14 points; modified G8: <6 points<sup>b</sup> G8: ≤14 points; modified G8: ≥6 points<sup>c</sup> Vulnerable or frail or too sick<sup>e</sup> Vulnerable or frail<sup>d</sup> Latent classes 2 to 4: Malnourished or cognitive/mood impaired or globally impaired



**Figure 5.1.** Receiver-operating characteristic (ROC) curves for predicting different reference standards: original vs. modified G8 questionnaire:

A) Geriatric Assessment  $\geq 1$  impairment; B) Geriatric Assessment  $\geq 2$  impairments; C)  $\geq 1$  geriatric intervention prescribed; D) SIOG classification (Fit vs. Vulnerable/Frail/Too sick); E) Balducci's classification (Fit vs. Vulnerable/Frail); F) Latent typology (fit vs. "unhealthy profiles").

**Table 5.6.** Diagnostic performances according to different reference standards (original vs. modified G8)

Reference standards	Prevalence	Screening tools	Cut-offs <sup>a</sup>	Sensitivity (95% CI)	p-value <sup>b</sup>	Specificity (95% CI)	p-value <sup>b</sup>	AUROC (95% CI)	p-value <sup>c</sup>
<b>GA <math>\geq 1</math> impairment</b>	91.9%	Original G8	$\leq 14$	90.7% (88.8%-92.4%)	0.921	54.3% (43.6%-64.8%)	<b>0.001</b>	0.90 (0.87-0.92)	<b>0.0029</b>
		Modified G8	$\geq 6$	90.6% (88.7%-92.3%)		75.0% (64.9%-83.4%)		0.93 (0.91-0.95)	
<b>GA <math>\geq 2</math> impairments</b>	76.9%	Original G8	$\leq 13.5$	88.7% (86.4%-90.7%)	<b>0.002</b>	58.4% (52.2%-64.4%)	<b>0.048</b>	0.87 (0.84-0.89)	<b>0.0006</b>
		Modified G8	$\geq 8$	91.9% (89.9%-93.6%)		64.8% (58.6%-70.5%)		0.90 (0.88-0.92)	
<b><math>\geq 1</math> geriatric intervention</b>	90.9%	Original G8	$\leq 14$	89.9% (87.9%-91.7%)	0.629	41.4% (31.8%-51.4%)	<b>0.005</b>	0.81 (0.77-0.86)	<b>0.0056</b>
		Modified G8	$\geq 6$	89.4% (87.4%-91.2%)		55.8% (45.7%-65.5%)		0.85 (0.81-0.89)	
<b>SIOG classification<sup>d</sup></b>	83.2%	Original G8	$\leq 13.5$	84.8% (82.3%-87.0%)	<b>0.006</b>	59.8% (52.1%-67.1%)	<b>0.016</b>	0.83 (0.80-0.86)	<b>&lt;0.00001</b>
		Modified G8	$\geq 8$	87.7% (85.5%-89.7%)		69.0% (61.5%-75.7%)		0.88 (0.86-0.91)	
<b>Balducci's classification<sup>e</sup></b>	86.5%	Original G8	$\leq 13.5$	83.2% (80.7%-85.5%)	0.241	56.9% (48.6%-64.8%)	0.835	0.80 (0.77-0.84)	0.2644
		Modified G8	$\geq 8$	84.5% (82.1%-86.7%)		57.5% (49.3%-65.5%)		0.82 (0.78-0.85)	
<b>Latent class typology<sup>f</sup></b>	79.5%	Original G8	$\leq 13.5$	88.2% (85.9%-90.2%)	0.999	62.2% (55.7%-68.5%)	0.103	0.86 (0.83-0.88)	0.1338
		Modified G8	$\geq 8$	88.2% (85.9%-90.2%)		57.1% (50.5%-63.5%)		0.84 (0.82-0.87)	

Abbreviations: AUROC, area under the ROC (receiver operating characteristic) curve; 95% CI, 95% confidence interval; GA, Geriatric Assessment; SIOG, International Society of Geriatric Oncology

<sup>a</sup> Official cut-off for GA  $\geq 1$  impairment and best cut-off (prioritizing sensitivity) otherwise.

<sup>b</sup> Original vs. modified G8 (McNemar's Chi-square test)

<sup>c</sup> Original vs. modified G8 (AUROC Chi-square test)

<sup>d</sup> Fit vs. Unfit (Vulnerable or frail or too sick)

<sup>e</sup> Fit vs. Unfit (Vulnerable or frail)

<sup>f</sup> Fit vs. Unfit (Latent classes 2 to 4: Malnourished or cognitive/mood impaired or globally impaired)

### 5.3.4 Sensitivity analyses

Sensitivity analyses produced closely similar results when patients with the specific reference definition available were also included in the analyses (Annex 7, **Table A7.2**).

When considered the IADL 4-items for men and 8-items for women in the definition of the SIOG classification, results were very similar as those using the 8-item IADL for all patients. For the original G8 sensitivity and specificity were 86.7% and 59.9%, respectively; for the modified G8, corresponding values were 89.6% and 67.8%. AUROCs were significantly higher for the modified G8: 0.90 (95% CI 0.87-0.92) vs. 0.85 (0.82-0.87) [original G8]; p-value <0.00001.

## 5.4 Discussion

### 5.4.1 Summary of findings

In the present study of elderly patients with cancer, we assessed the diagnostic performance of the original and modified G8 tools in different contexts in order to evaluate their robustness. Six definitions of reference standards that evocate a geriatric risk profile were tested. Regardless of the definition tested, both tools demonstrated high predictive value and performance robustness to detect frailty. Comparing the original and modified G8, statistically significant differences were found in favor of the modified G8 between AUROC for GA  $\geq 1$  and  $\geq 2$  impairments (p=0.0029 and 0.0006 respectively), a major geriatric intervention (p=0.0056), and the SIOG classification (p<0.00001), demonstrating better screening performances of the modified G8.

Interestingly, both G8 and modified G8 were found to be predictive of the subsequent prescription of geriatric interventions on relevant clinical domains. This finding is of particular clinical relevance, as it relates directly to the main objective of the screening tools to identify those patients who would benefit from a complete GA. Beyond conceptual pitfalls to define frailty, this further supports the pragmatic aim of the G8 instruments to provide adequate detection of patients with potential deficits warranting interventions and optimization of the treatment cancer management.

G8 and modified G8 screening tools were originally developed to identify patients with at least one impairment in a multidimensional GA, which has been proposed by the SIOG [27] as the reference standard for evaluation of the elderly cancer patient to determine the optimal oncologic treatment. However, a standardized definition of GA and, more importantly, abnormal GA is lacking. Indeed, the definition of what is considered to be an abnormal GA varies largely across studies, which may use different number of components and different scales and thresholds for defining impairment, hence limiting comparability of study results [56]. Furthermore, this pragmatic definition most often used in the literature does not correspond well to the reality of clinical practice of geriatricians and oncologists, having limited applicability and representing a problem for implementation in routine clinical care.

Other frailty classifications have been developed in order to assist physicians to select the best cancer treatment and guide geriatric interventions. In a recent study, the prognostic value of three of those classifications (Balducci, SIOG and Ferrat's LCT) was assessed and found to be good for one-year mortality and six-months unscheduled hospitalizations in older patients with cancer [149]. This supports their use to stratify older cancer patients according to their health status for clinical decision making

process, and also as a candidate reference definition for screening test accuracy studies because of their predictive value for patient outcomes [147]. Some frailty criteria for example have been used to help with evaluation of treatment toxic effects.

#### 5.4.2 Comparison with other reviews

The original G8 has been compared with GA in 16 studies [61,67,68,70,71,73-75,85,86,88,120-122,153,154] in older patients with cancer, of which 7 studies used a cutoff for impairment of  $\geq 1$  deficiency at GA, reporting sensitivity ranging from 65% to 90% (91% in our study) and specificity ranging from 3% to 100% with an average of 55% (54% in our study), and 12 studies also reported results using a cutoff for impairment of  $\geq 2$  deficiencies at GA, with sensitivity ranging from 38% to 97% (95% in our study) and specificity from 29% to 79% (40% in our study). Another study [70] of patients with hematologic disorders used Fried's criteria to assess the performance of the G8, reporting results of similar magnitude with a sensitivity of 82% and a specificity of 51%. Outside the oncological setting, only few studies evaluated screening tools against definitions others than GA [155,156]. To our knowledge, no other reference standard was tested for the G8, and the present analysis is the first study to report on the diagnostic performance of the modified G8 using gold standards others than an abnormal GA.

#### 5.4.3 Strengths and limitations

The present study is the first to thoroughly examine the variability of the diagnostic performance of screening tools for frailty in older patients with cancer under multiple clinically relevant reference definitions. Adding to the previously reported high

prognostic value of the two instruments [157], those findings reinforce the clinical utility of the G8 tools in daily geriatric oncologic practice.

Our study has limitations that should be noted. First, data were missing for some key variables to compute G8 scores and/or reference standards, although missing rates per variable was overall low (median 7%, range 0%-17.6%). Relatedly, patients were excluded from the present analysis when data on any of the six reference definitions was not available to allow direct comparison of the performance of the screening tools under varying reference standards using a common population. Of note, no statistically significant difference was found between included and excluded patients regarding main demographic and clinical characteristics (Annex 3, **Table A3.3**), and the similar results of the sensitivity analyses support the robustness of our findings. It would have also been of interest to assess other approaches as reference standards, such as the Fried phenotype [22] and the Rockwood's frailty index [158], two well-established instruments measuring frailty but developed for the general geriatric population and not specifically for older patients with cancer.

## 5.5 Conclusion

Our findings demonstrate the robustness of the original and modified G8 to modifications of the reference gold standard, with evidence of a better diagnostic performance of the modified G8 for detecting a variety of health profiles evocative of frailty. These results further support the clinical value of these instruments for detecting older patients with cancer warranting a complete geriatric assessment.

## CHAPTER 6

---

### **Prognostic value of the original and modified G8 screening tools**

## **6.1 Introduction**

As a result of longer life expectancy, the proportion of elderly people among patients with cancer is growing and great efforts have been invested in developing more targeted care to older patients with cancer. Establishing clear prognoses and making optimal treatment decisions are crucial but challenging tasks due to the heterogeneity of the elderly population. There is a need to identify vulnerable patients at higher risk of poor outcomes and who would benefit from specific interventions and/or treatment adaptation [11,12].

The G8 and modified G8 screening tools were developed to help identify those vulnerable older patients with cancer who need a complete GA, which is time- and resource-consuming. An additional desirable property of a screening tool is its prognostic ability to predict further outcomes such as survival or functional decline.

The G8 was derived from the mini-nutritional assessment (MNA) because of its high predictive value for survival [61,85], but only a handful of studies have so far reported results on its prognostic value for survival, with limited information by tumor site despite evidence for possible heterogeneity in this regards [56,57,67].

The aim of this study was consequently to assess and compare the prognostic value of the original G8 to its optimized 6-item version (modified G8) in a large cohort of older patients with cancer, overall and by tumor site.

## 6.2 Methods

### 6.2.1 Study population

The ELCAPA prospective cohort is described in Chapter 2 (General methodology). **Figure 2.3** presents the flow chart of participants included in this study. Because the main objective of the present analysis was to provide direct comparisons of the prognostic values of the two G8 versions, patients with missing G8 and/or modified G8 were excluded from analysis to ensure an identical population for assessing both instruments. Results were found to be closely similar when patients with missing modified G8 but available G8 were also included in the analyses (Annex 8, **Table A8.1**)

### 6.2.2 Endpoints

Endpoints were overall 1- and 3-years survival, defined as the time from evaluation to death within 1 and 3 years or to the last follow-up for censored patients: 1) patients alive at analysis cutoff time points, and 2) those lost to follow-up before the analysis cutoff time points. All patients had a minimum follow-up of one year.

### 6.2.3 Screening tools

The 2 screening tools under study are described in detail in **Table 1.6** (original G8) and **Table 3.4** (modified G8). Briefly, both tools include items relating to medication, nutritional, cognitive and functional status. The 8-item G8 score ranges from 0 to 17, a *higher* score indicating better health status (Abnormal G8 score:  $\leq 14$  [61]) and the 6-item modified G8 score ranges from 0 to 35, a *lower* score indicating better health status (Abnormal modified G8 score:  $\geq 6$  [122]).

#### 6.2.4 Statistical Analysis

The endpoints 1- and 3-years overall survival were calculated by the Kaplan–Meier method and compared across groups by the log-rank test. Crude hazard ratios (HRs) with corresponding 95% confidence intervals (CIs) were first estimated by unadjusted Cox proportional-hazards analysis, considering an abnormal G8 or a modified G8 score as the exposure variable.

Multivariate analysis was then conducted after adjusting for age, gender, tumor site, metastatic status and anticancer treatment (surgery, chemotherapy, radiotherapy, targeted therapy and hormone therapy). We tested the interaction term between tumor site and metastatic status, in light of a previously reported finding from the ELCAPA cohort [32]. Associations were evaluated in the whole study population and after stratification by metastatic status (excluding hematological malignancies) and tumor site to test the robustness of the results in varying clinical situations.

To further assess the prognostic value of the screening tools across their respective ranges, complementary analyses were performed after categorizing continuous scores into classes of increasing risk. Recursive partitioning analysis (RPA) was carried out by recursive partitioning of overall survival to determine the optimal thresholds to separate the classes. This methodology used martingale residuals from a Cox model to determine the optimal value among all possible cut-points for dividing the data into 2 sets with the greatest difference in outcome. The procedure was then repeated in resulting groups until predefined stopping criteria were met.  $P < 0.05$  was used for this analysis, with a minimal resulting group size of 100 and until a maximum of 5 classes was identified. This analyses involved an implementation of RPA for Stata by Wim van Putten [159].

Finally, the prognostic value of the ECOG-PS and the TUG, two short instruments commonly used in geriatric oncology, was evaluated and compared with that obtained from the G8 and modified G8.

Discriminative performance of all models was evaluated by Harrell's C-index [160] and Gönen and Heller's K concordance index [161]. Harrell's C-index is defined as the proportion of all patient pairs in which the predictions and outcomes are concordant. Because of censoring, not all pairs are evaluable. Gönen and Heller proposed an alternative estimator to avoid bias due to censoring. It involves only the regression parameters and the covariate distribution and is therefore asymptotically unbiased.

The proportional-hazards assumption was tested by using Schoenfeld residuals and retained. For covariates perfectly predicting survival (e.g. when all deceased patients have an abnormal score), a Firth's penalization procedure was applied to compute the HRs.

## 6.3 Results

### 6.3.1. Study population

Between January 2007 and April 2014, 1613 patients were recruited for the ELCAPA cohort. For the present study, we used data for 1333 patients with complete follow-up and data for the G8 and modified G8 available at the time of analysis (**Figure 2.3**). Compared to the population not included in the analysis (n=280), included patients were younger (median age=80 years vs. 81, p=0.0002), had better score G8 (median=11 vs. 10, p=0.003) and a higher proportion of patients had a better performance status (PS 0-1: 49.4% vs. 42.3%, p=0.002). There were more patients with prostate cancer and fewer patients with upper gastrointestinal cancer in the study population (p=0.003) (**Annex 3, Table A3.4**).

Baseline characteristics of the study population were as follows: median age was 80 years, and 51.8% were men. The most common tumor sites were colorectal (19.6%), upper gastrointestinal tract and liver (17.1%) and breast (16.4%). Cancers were in metastatic stage for 49.3% patients. Prevalence of abnormal GA, as defined by at least one abnormal test score, was 92.0% (n=1170), and G8 and modified G8 scores were abnormal for 83.6% (n=1115) and 83.1% (n=1108) of patients, respectively.

Treatment modalities by order of decreasing frequency were chemotherapy (34%), surgery (22%), radiotherapy (20%), hormone therapy (14%) and targeted therapy (2%). Two hundred and forty-five patients (18%) did not receive any treatment.

### 6.3.2. Survival analyses in the whole study population

For evaluating overall survival, the median follow-up time was 26.5 months (range 0.03–92.7) from the initial evaluation. The median survival was 17.6 months [IQR 4.1–66.8], and the 1- and 3-years overall survival were 58.4% [95% CI 55.4%-60.8%] and 36% [33.1%-38.9%], respectively.

A significant difference in median survival was found between normal and abnormal scores for both tests. Considering G8 score and its validated cutoff value, the median survival was 76 months [IQR 27-not applicable] for patients with a normal G8 score (>14/17) and 13.1 months [3.3-42.6] for those with an abnormal score (p-value <0.0001). Similar results were found for the modified G8, with a median survival of 76 months [21-not applicable] for patients with a normal score (<6/35) and 13.1 [3.3-45.9] with an abnormal score (p-value <0.0001).

Kaplan–Meier curves for overall 3-years survival by G8 and modified G8 scores are shown in **Figure 6.1**. Patients with normal and abnormal scores showed significant differences for both screening tools (log-rank  $P < 0.0001$ ). After dividing the continuous scores into classes of increasing risk by RPA (n=5 classes identified for each score), we observed a clear graded relationship between classes with worsening scores and 1- and 3-years survival (both global log-rank  $P < 0.0001$ ).

Results from unadjusted and adjusted Cox proportional-hazards analyses in the whole study population are in **Table 6.1**. On multivariate analysis, abnormal scores for both scales were independently associated with overall 1-year survival (G8: adjusted HR [aHR]=4.31 [95% CI 2.73-6.80],  $p < 0.0001$ ; modified G8: aHR=4.87 [3.10-7.64],  $p < 0.0001$ ) and 3-years survival (G8: aHR=2.94 [2.17-3.98],  $p < 0.0001$ ; modified G8 aHR=2.56 [1.95-3.37],  $p < 0.0001$ ).

We found a statistically significant interaction between tumor site and metastatic disease ( $p \leq 0.0001$ ), showing increased mortality in patients with metastatic prostate cancer (e.g. HR for overall 3-years survival=13.11 [95% CI 6.90-24.93]) and breast cancer (HR=6.84 [3.96-11.84]), while the HRs for other tumor sites ranged from 2 to 3.5. To test the stability of the results, we created a model including an interaction term between metastatic status and tumor site and found results of similar magnitude (abnormal G8 aHR=2.56 [1.88-3.47];  $p < 0.0001$ ; abnormal modified G8 aHR=2.36 [1.79-3.11];  $p < 0.0001$ ). Unadjusted and adjusted analyses of the categorical classes revealed similar results as those for Kaplan–Meier estimates, with progressively increasing mortality risk with worsening scores, also indicating higher performance indices than those obtained using binary abnormal scores (e.g. unadjusted C-indexes for modified G8 0.58 [binary] vs. 0.69 [categorical]).

Results regarding the prognostic value of the ECOG-PS and the TUG are shown in Annex 8 (**Table A8.2**). An increasing ECOG-PS and, to a lesser extent, an abnormal TUG ( $\geq 20$ s) were both significantly associated with poorer 1- and 3-years survival. In particular, results for the ECOG-PS closely matched those obtained with the G8 and modified G8 when using classes of increasing risk, as indicated by the very similar prognostic performance indices and range in Hazard Ratios.

### 6.3.3. Survival analyses by metastatic status and cancer sites

Complementary analyses by metastatic status are shown in **Figure 6.2** (Kaplan–Meier curves) and **Table 6.2** (Cox proportional-hazards models). Overall 3-years survival rates differed by G8 and modified G8 score [without metastases, modified G8: 81% (normal score) vs. 50% (abnormal score); G8: 84% (normal score) vs. 48% (abnormal score); with metastasis, modified G8: 34% (normal score) vs. 12% (abnormal

score); G8: 37% (normal score) vs. 12% (abnormal score)]. Considering the associated HRs, abnormal scores highly predicted poor outcome, regardless of the presence or absence of metastasis and for both screening tools.

In line with our initial objective, we analyzed both scores by tumor site (**Figure 6.3**). Unadjusted and adjusted associations between abnormal scores and overall 1-year survival were statistically significant for most sites, with the exception of the upper digestive tract and liver for both tools (G8 aHR=1.55 [95% CI 0.68-3.6]; p=0.297; modified G8 aHR=2.24 [0.91-5.51]; p=0.081) and colorectal cancer for G8 only (aHR=2.38 [0.85-6.68], p=0.100; modified G8 aHR=3.50 [1.09-11.25], p=0.035). For hematological malignancies and breast cancers, all deceased patients had abnormal G8 and modified G8 scores, which prevented the direct calculation of the HRs (infinite values). Cox analysis with Firth's penalization procedure yielded the following results: hematological malignancies: HR=12.2 [G8]/16.4 [modified G8], aHR=10/12.8; breast cancer: HR=34.6/10.3, aHR=15.5/20.3; all p<0.001. Similarly, both abnormal scores were significantly associated with overall 3-years survival for all but digestive cancers. For hematological malignancies results after Firth's procedure were as follows: HR=19.8 [G8]/8.4 [modified G8]; aHR=17.1/6.5; all p<0.001. Both instruments had statistically significant associations with survival in all tumor sites when using classes of increasing risk (Annex 8, **Table A8.3**).

**Tables A8.4** and **A8.5** (Annex 8) show discrimination performance indices by tumor site, indicating substantially higher indices when using scores in classes of increasing risk: e.g. C-indexes from 1-year adjusted analyses of abnormal scores ranged from 0.69 (hematological [G8] and other tumor sites [both tools]) to 0.86 (prostate cancer [modified G8]), while analyses of risk classes ranged from 0.72 (urinary tract cancer [modified G8]) to 0.89 (breast and prostate cancer [modified G8]).

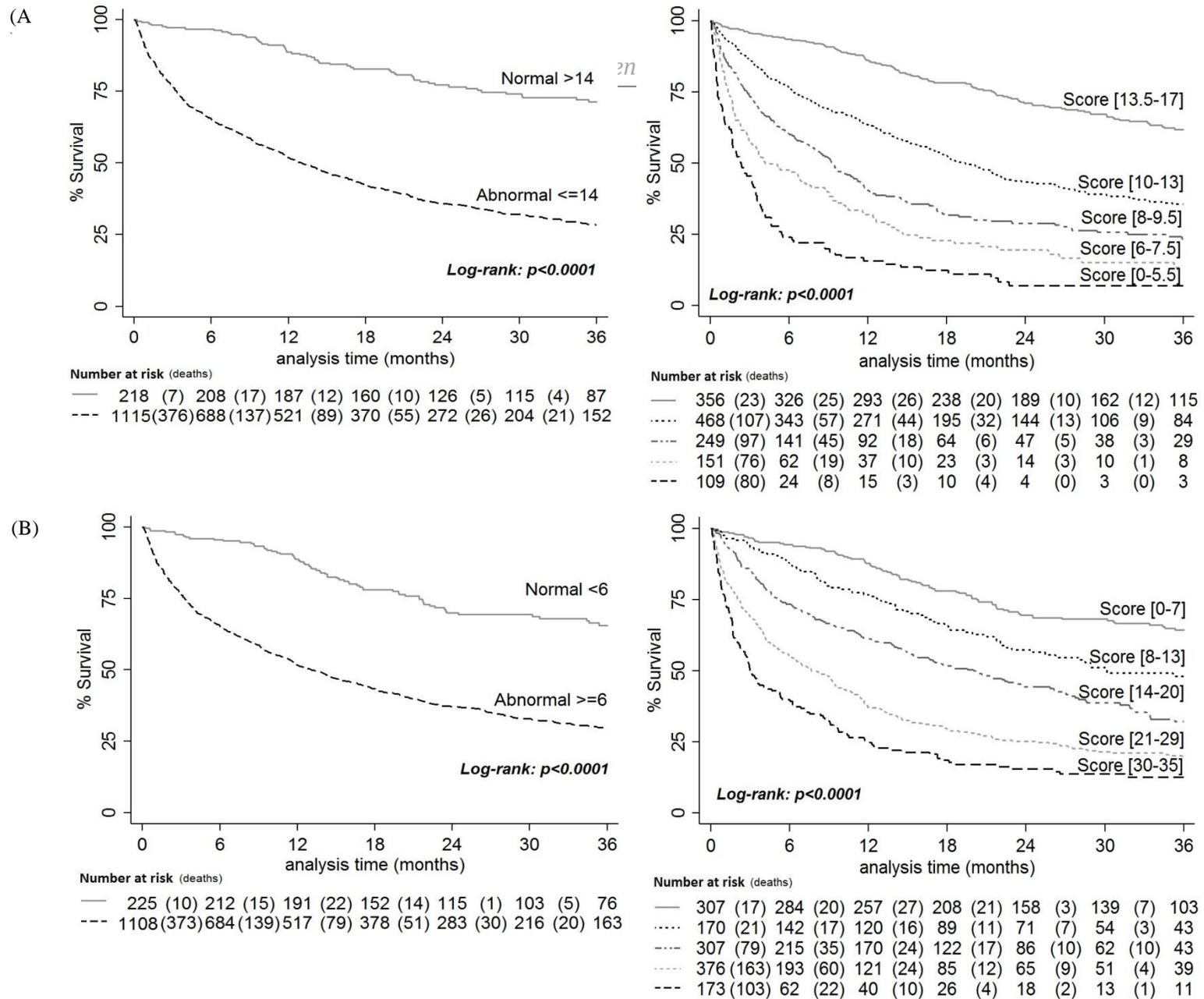


Figure 6.1. Kaplan-Meier curves of overall 3-years survival for G8 (A) and modified G8 (B) scores

**Table 6.1.** Comparison of the prognostic value for overall 1- and 3-years survival of the screening tools G8 and modified G8: unadjusted and adjusted Cox proportional hazards models

Outcome	Screening tools	Unadjusted analysis N=1333								Adjusted analysis N=1168					
		N (%)	Deaths (%)	HR	95% CI	p-value	Harrell-C	Gönen <sup>a</sup>	N (%)	aHR	95% CI	p-value	Harrell-C	Gönen <sup>a</sup>	
1-year survival	G8	Abnormal ≤14	1115 (84)	513 (95)	5.7	3.8-8.5	<0.0001	0.58	0.60	979 (84)	4.3	2.7-6.8	<0.0001	0.74	0.73
		13.5-17	356 (27)	48 (13)	1(ref)		<0.0001	0.71	0.70	312 (27)	1(ref)		<0.0001	0.79	0.77
		10-13	468 (35)	164 (35)	3.1	2.2-4.2				416 (36)	2.7	1.9-3.9			
		8-9.5	249 (19)	142 (57)	6.0	4.3-8.4				216 (18)	4.5	3.1-6.5			
		6-7.5	151 (11)	95 (63)	8.4	5.9-11.9				129 (11)	5.7	3.8-8.4			
		0-5.5	109 (8)	88 (81)	14.8	10.3-21.0				95 (8)	10.3	6.8-15.4			
	Modified G8	Abnormal ≥6	1108 (83)	512 (95)	5.6	3.8-8.4	<0.0001	0.58	0.60	971 (83)	4.9	3.1-7.6	<0.0001	0.75	0.73
		0-7	307 (23)	37 (12)	1(ref)		<0.0001	0.72	0.70	272 (23)	1(ref)		<0.0001	0.78	0.75
		8-13	170 (13)	38 (22)	2.0	1.3-3.1				148 (13)	1.9	1.2-3.0			
		14-20	307 (23)	114 (37)	3.8	2.6-5.5				264 (23)	3.2	2.1-4.8			
		21-29	376 (28)	223 (59)	7.6	5.4-10.8				335 (29)	5.9	4.1-8.7			
		30-35	173 (13)	125 (72)	11.6	8.1-16.8				149 (13)	8.9	5.9-13.3			
3-years survival	G8	Abnormal ≤14	1115 (84)	704 (93)	4.0	3.1-5.3	<0.0001	0.58	0.58	979 (84)	2.9	2.2-4.0	<0.0001	0.75	0.73
		13.5-17	356 (27)	116 (33)	1(ref)		<0.0001	0.69	0.67	312 (27)	1(ref)		<0.0001	0.77	0.74
		10-13	468 (35)	262 (56)	2.3	1.8-2.9				416 (36)	1.9	1.5-2.5			
		8-9.5	249 (19)	174 (70)	3.7	2.9-4.7				216 (18)	2.7	2.1-3.5			
		6-7.5	151 (11)	112 (74)	5.5	4.2-7.1				129 (11)	3.8	2.8-5.1			
		0-5.5	109 (8)	95 (87)	9.3	7.1-12.3				95 (8)	6.8	5.0-9.4			
	Modified G8	Abnormal ≥6	1108 (83)	692 (91)	3.2	2.5-4.1	<0.0001	0.58	0.57	971 (83)	2.6	1.9-3.4	<0.0001	0.75	0.73
		0-7	307 (23)	95 (31)	1(ref)		<0.0001	0.69	0.66	272 (23)	1(ref)		<0.0001	0.77	0.74
		8-13	170 (13)	75 (44)	1.6	1.2-2.2				148 (13)	1.5	1.1-2.0			
		14-20	307 (23)	175 (57)	2.5	2.0-3.3				264 (23)	1.9	1.4-2.5			
		21-29	376 (28)	272 (72)	4.4	3.5-5.6				335 (29)	3.0	2.3-3.9			
		30-35	173 (13)	142 (82)	6.7	5.1-8.7				149 (13)	4.5	3.4-6.1			

Abbreviations: HR, unadjusted hazard ratio; HRa, hazard ratio adjusted for age, gender, anticancer treatment, cancer site and metastasis; 95% CI, 95% confidence interval.

<sup>a</sup> Gönen and Heller's K concordance statistic

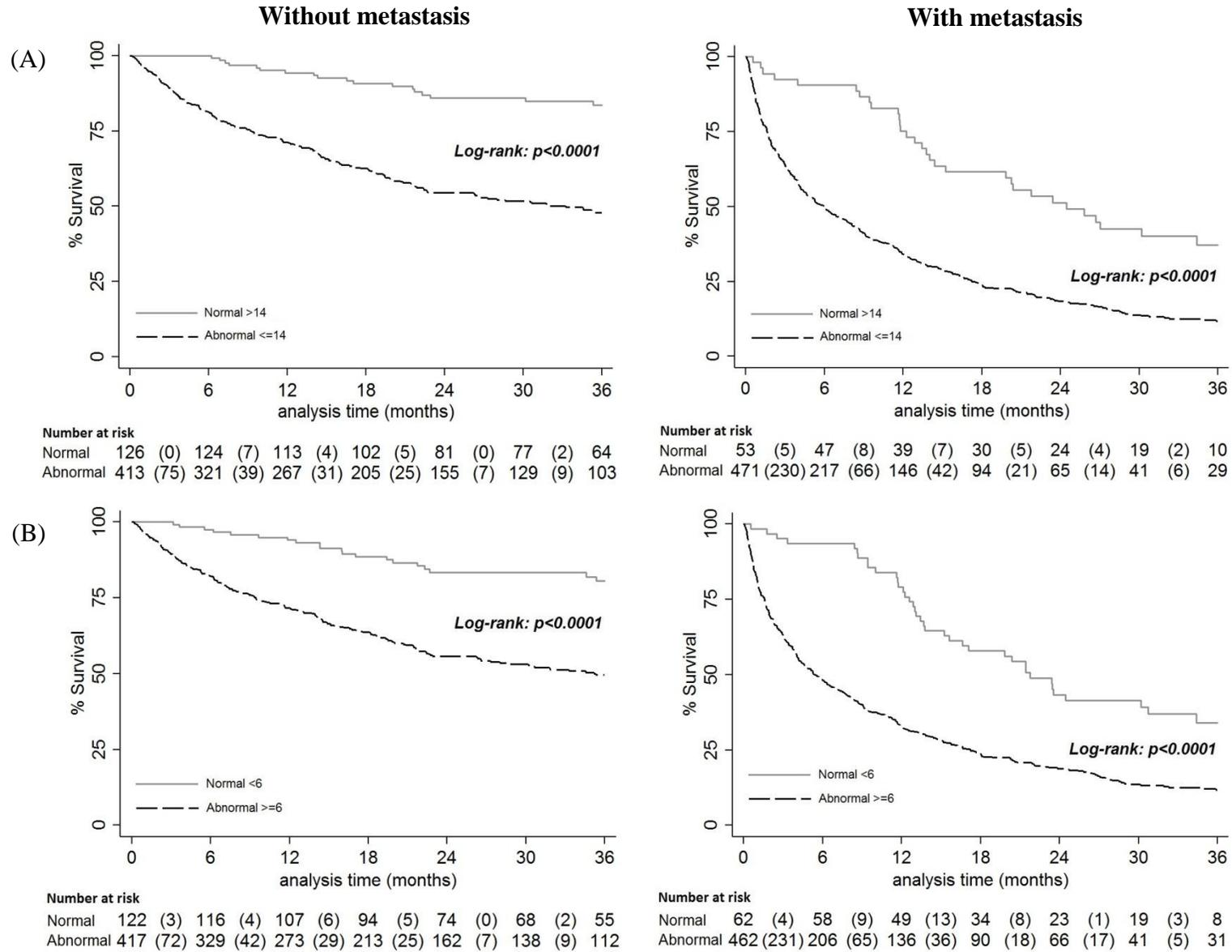


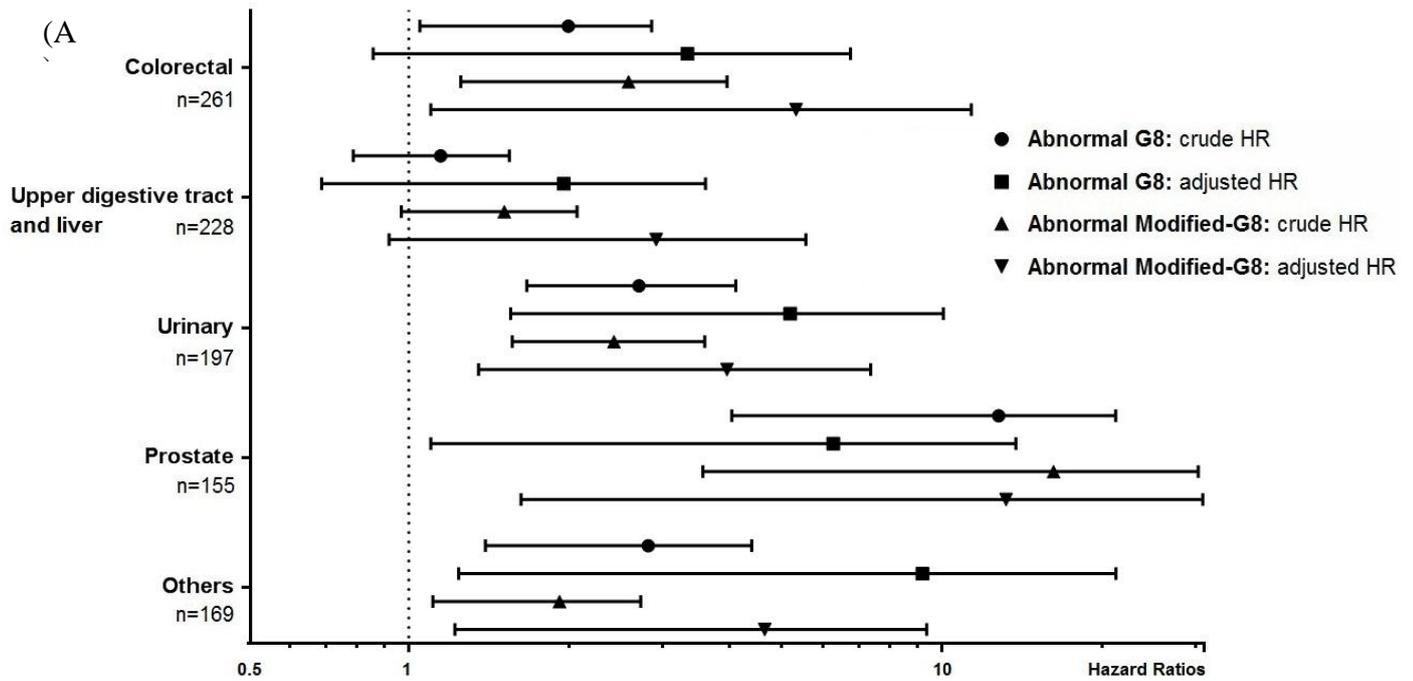
Figure 6.2. Kaplan-Meier curves of overall 3-years survival for G8 (A) and modified G8 (B) scores, stratified by metastatic status

**Table 6.2.** Comparison of the prognostic value for overall 1- and 3-years survival of the screening tools G8 and modified G8, stratified by metastatic status: unadjusted and adjusted Cox proportional hazards models

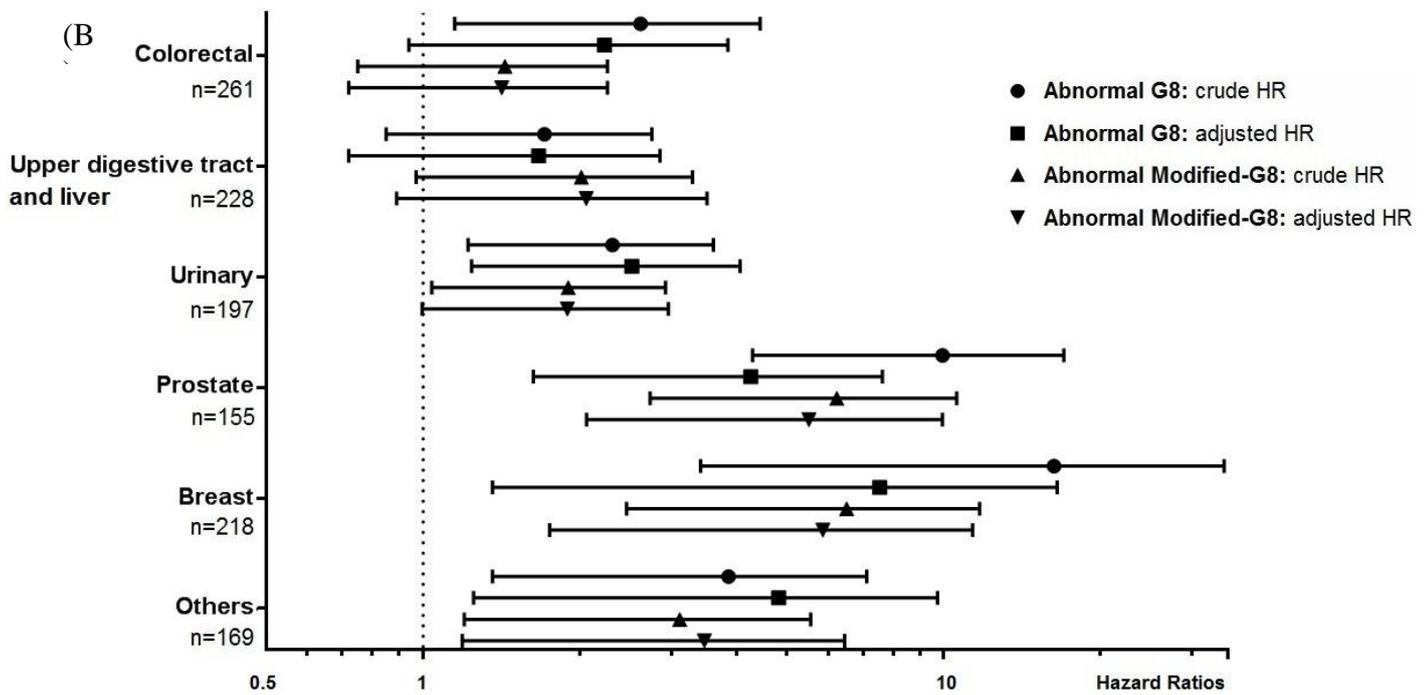
Outcome	Metastatic status		Unadjusted analysis							Adjusted analysis					
			N (%)	Deaths (%)	HR	95%CI	p-value	Harrell-C	Gönen <sup>a</sup>	aHR	95%CI	p-value	Harrell-C	Gönen <sup>a</sup>	
1-year survival	With metastasis, n=524					N=1063						N=1063			
	<i>G8 score</i>	Abnormal ≤14	471	296 (95.8)	3.99	2.29-6.96	<0.0001	0.55	0.55	3.92	2.23-6.88	<0.0001	0.65	0.64	
	<i>Modified G8 score</i>	Abnormal ≥6	462	296 (95.8)	5.03	2.88-8.77	<0.0001	0.57	0.57	4.73	2.71-8.26	<0.0001	0.65	0.65	
	Without metastasis, n=539														
	<i>G8 score</i>	Abnormal ≤14	413	114 (94.2)	5.93	2.77-12.73	<0.0001	0.61	0.63	2.68	1.18-6.06	0.019	0.76	0.76	
	<i>Modified G8 score</i>	Abnormal ≥6	417	114 (94.2)	5.54	2.58-11.88	<0.0001	0.60	0.62	2.90	1.31-6.39	0.008	0.77	0.76	
3-years survival	With metastasis, n=524														
	<i>G8 score</i>	Abnormal ≤14	471	379 (92.4)	2.58	1.79-3.73	<0.0001	0.55	0.54	2.31	1.59-3.35	<0.0001	0.66	0.63	
	<i>Modified G8 score</i>	Abnormal ≥6	462	372 (90.7)	2.48	1.78-3.47	<0.0001	0.56	0.54	2.24	1.60-3.16	<0.0001	0.67	0.63	
	Without metastasis, n=539														
	<i>G8 score</i>	Abnormal ≤14	413	186 (91.2)	4.37	2.69-7.09	<0.0001	0.61	0.61	2.76	1.65-4.64	<0.0001	0.73	0.74	
	<i>Modified G8 score</i>	Abnormal ≥6	417	184 (90.2)	3.50	2.21-5.56	<0.0001	0.59	0.60	2.21	1.36-3.59	0.001	0.72	0.73	

Abbreviations: HR, unadjusted hazard ratio; aHR, HR adjusted for age, gender, anticancer treatment and cancer site; 95% CI, 95% confidence interval.

<sup>a</sup> Gönen and Heller's K concordance statistic



Note: perfect survival predictions for breast cancer and hematological malignances (all deceased patients had an abnormal score)



Note: perfect survival predictions for hematological malignances (all deceased patients had an abnormal score)

**Figure 6.3.** Prognostic value for overall A) 1- and B) 3-years survival of G8 and modified G8 scores according to tumor sites: unadjusted and adjusted Cox proportional-hazards models

HR, Hazard Ratio; adjusted HR, Hazard Ratio adjusted for age, gender, treatment and metastatic status.

Others: Unknown primary origin (n=31), lung (n=36), skin (n=46), sarcoma (n=15), gynecologic (n=13), brain (n=9), head and neck (n=5), thyroid (n=4), others (n=10).

## 6.4 Discussion

### 6.4.1 Summary of findings

The aim of the present study was to assess the prognostic value of the modified 6-item G8 and to compare it to that of the original G8 in a large cohort of older patients with cancer. Abnormal scores on both instruments independently and highly predicted overall 1-year survival, with a graded relationship between worsening scores and increased mortality. This prognostic performance persisted for both instruments after stratifying by tumor site and regardless of metastasis status.

### 6.4.2 Comparison with other reviews

We found poor overall 1-year survival among patients with a geriatric risk profile, as indicated by an abnormal G8 or modified G8 score. This finding remained after adjusting for age, gender, tumor site and metastatic status, which reinforces the independent prognostic value of both screening tools. Our finding of a high prognostic value for the original G8 is consistent with previous studies, which reported statistically significant HRs for overall survival when considering an abnormal versus normal score in older patients for a variety of tumor sites (HR=2.63 [95% CI 1.92-3.70] in Kenis et al. [88]; HR=4.72 [3.07-7.26] in Soubeyran et al. [85]). This positive association might be linked to the development process of the G8, which was initially elaborated from the MNA short form (MNA-SF) because of its prognostic value in older patients [95,118] and thus incorporates 7 MNA-SF items [61]. Alternatively, the modified G8 was developed to optimize the ability of the G8 to discriminate between fit and vulnerable patients who would benefit from a GA, as indicated by the detection of at least one abnormal test

result during the complete GA. As such, this modification was specifically targeted at strengthening the properties of the G8 as a screening instrument, with no consideration of the prognostic value of the resulting tool. Despite this difference and a lower number of items (6 vs 8), our findings demonstrate a persistent high predictive value for survival with the modified G8.

Similar to its original version, the modified G8 covers multiple domains assessed during a complete GA, including nutritional aspects, mood, cognition, mobility, polypharmacy and self-rated health status. The prognostic value of alterations in such domains have previously been demonstrated in older patients with various cancer types [67,94,117,118], with evidence for consistent associations of malnutrition, comorbidity and functional status with mortality [32]. Self-rated health status has been shown to significantly predict mortality in older populations [162-164]. In comparison to the original G8, the modified G8 also includes a simplified version of the ECOG-PS, which is commonly used in oncology settings and whose score was found independently associated with poor survival in several studies [85,165,166]. In our study, the ECOG-PS assessed separately was also found to be strongly and independently associated with overall survival, with prognostic performance indices remarkably similar to those obtained from the G8 instruments. An abnormal TUG ( $\geq 20$ s), another short instrument of common use in oncological practice, also demonstrated prognostic value for survival, though with slightly inferior performance indices, but consistently with previous reports [167]. While the multi-item G8 instruments did not strictly outperform the ECOG-PS in terms of predictive power for survival, it should be stressed that the former were specifically designed as screening tools for detection of vulnerable patients in need of a comprehensive GA, an endpoint for which the ECOG-PS has proven inconsistently but generally poorly predictive. One study conducted in 135 patients aged  $\geq 65$  years with

solid abdominal tumors in need of surgery thus reported a sensitivity of 54% and a specificity of 81% at cut-off  $\geq 2$  for identifying  $\geq 2$  geriatric abnormalities on GA [73], while another study in 117 patients aged  $\geq 65$  years with various cancer types found corresponding values of 94% and 55% at cut-off  $\geq 1$  [168]. To our knowledge, the diagnostic properties of the TUG have never been evaluated in older patients with cancer.

Our study also identified a graded association between worsening scores and increasing risk of mortality for both G8 screening tools. We determined 4 to 5 classes of increasing risk for each instrument, to overcome the simplistic conclusion of normality or abnormality. By providing differentiated and more accurate prognostic information, our findings reinforce the clinical utility of these screening instruments, as indicated by the substantially increased predictive performance of both instruments (Harrell and Gönen indices). This information is of particular interest for clinical practice in the geriatric oncology setting, where decisions and modalities of treatment for older patients with cancer often rely on personal experience and clinical judgment [169,170]. Our study strengthens the utility of both the original and modified G8 as short and easy tools for the oncologist and geriatrician to evaluate prognosis and the further need for a complete GA. The latter tool still remains invaluable to identify previously undetected geriatric problems for which targeted interventions can be applied [28].

We analyzed the data by tumor site to examine the stability of the results across a variety of cancer types, including colorectal, upper digestive tract and liver, breast and urinary and prostate cancers as well as hematological malignancies. Even after multiple adjustments, both instruments were prognostic for survival in all sites using classes of increasing risk and most sites using (ab)normal thresholds (6 of 7 sites for the modified

G8 and 5 of 7 for the original G8). Exceptions included upper digestive tract and liver cancers for the G8 and modified G8, and colorectal cancers for the G8 only in the 1-year survival analysis and for both tools in the 3-years survival analysis. Again, prognostic performance indices noticeably increased when considering classes of increasing risk. Data remain scarce regarding the prognostic value of the G8 by tumor type. To our knowledge, our study is the first to contrast such prognostic results by cancer type in a unified cohort population. Two previous studies investigated older populations with various cancers but did not report prognostic information by subgroups [85,88]. One study of patients with head and neck cancer found a significant association between an abnormal G8 score and survival (aHR=3.19 [95% CI 1.48-6.87][171]). Two other studies of patients with haematological malignancies reported mixed results: Hamaker et al [67] found the G8 score an independent predictor of survival (HR 3.93 [1.67 – 9.22]), whereas Dubruille et al [172] found no significant association. Our findings are consistent with those from Hamaker et al., a result that probably reflects the shared characteristics of the populations for both studies (i.e., patients with high prevalence of geriatric conditions: >90% of patients with at least one impairment in a geriatric domain), which confirms the clinical interest of the G8 and modified G8 in this population. In contrast, Dubruille et al. investigated a more selected population of 85 older patients deemed sufficiently fit to receive chemotherapy and found that neither the G8 nor the GA components predicted overall survival. Despite good performance and significant associations with survival when using classes of increasing risk instead of binary thresholds, an overall slightly lower prognostic value of both instruments was found in patients with digestive cancer, suggesting room for further improvement, e.g. by identifying features more specific to those patients with particularly poor prognosis. Of note in our study, those patients had more frequently an abnormal GA (98% of  $\geq 1$

impairment on GA vs. 89% in non-digestive cancers), worse performance status, poor nutritional status (Annex 8, **Table A8.6**) and survival (**Figure A8.1**), and substantially high surgery rates (colorectal: 36%; upper digestive tract: 17%), when surgery has been associated with poorer outcomes over the first year in patients with digestive cancers [173,174]. The G8/modified G8 tools were initially developed in patients with a variety of cancer types, by retaining variables demonstrating good overall diagnostic properties for use in a population of older patients with cancer, regardless of its localization. As a consequence, it is likely those instruments may have slightly less discriminative power when evaluating prognosis in such patients with particularly poor features and outcomes, also possibly overlooking the added impact of subsequent surgery.

Comparison of the prognostic value of the G8 instruments with those of other screening tools remains difficult because of the quasi-absence of direct comparisons between instruments in common populations. Available tools commonly used to identify patients likely to benefit from a complete GA include the Vulnerable Elders Survey-13 (VES-13) [79], the Groningen Frailty Indicator (GFI) [120] and the Flemish version of the Triage Risk Screening Tool (fTRST) [77]. Kenis et al. [88] directly compared the G8 to the fTRST using 2 different thresholds ( $\geq 1$  fTRST(1) and  $\geq 2$  fTRST(2)) and found a stronger predictive value for overall survival with the G8. Other studies did not compare with the G8 for survival and were mostly based on small series of patients. With the VES-13, a significant association with overall survival was reported in 77 patients with cancer who were  $\geq 60$  years old (HR=1.14; p=0.005) [72]; changes in VES-13 were associated with overall survival during chemotherapy in 21 older patients with digestive cancer (HR=1.24 [95% CI 1.05–1.48]) [93]. With the GFI, an abnormal GFI score ( $\geq 4$ ) predicted survival in 202 older patients with various cancers (HR=1.80 [1.17–2.78]) [95], and in 55 breast cancer patients (HR 3.46 [1.69 to 7.10]) [94]. No significant

association with short-term (30-day) mortality was found with the VES-13 or the GFI in a series of 22 older patients undergoing surgery for digestive cancer [175].

#### 6.4.3 Strengths and limitations

Among the strengths of the present study are its large sample size from a multicenter prospective cohort of older patients with cancer and the multiple sensitivity analyses performed to assess the stability of the results. To our knowledge, this is the first report of the prognostic value of the modified G8 and the first to assess the variability of the predictive performance of screening tools across tumor sites, by metastatic status or using classes of increasing risk. The limitations include the lack of information regarding outcomes other than overall survival, such as functional decline or treatment-related toxicity. Second, data for other screening tools were not available to allow for direct comparisons. Finally, our findings relate by design to a specific population characterized by a high prevalence of geriatric syndromes, which may question the generalizability of our results. However, the estimated 83% of patients with an abnormal G8 score is compatible to findings in other studies of the instrument, as reported by a recent review by Decoster et al. (median 76% [IQR 68-80] in 7 studies)[56].

Further research is required to provide more direct comparisons between tools in various settings. Yet, potential advantages for using the G8 or the modified G8 may stem from the fact that they have been specifically developed for older patients with cancer [61]. From our results, their prognostic value appears both high and consistent across tumor sites and regardless of metastatic status. Combined with their shortness and

simplicity of use, the G8/modified G8 screening tools can provide valuable clinical insights for geriatric oncology practice; yet, they should not be seen as substitution for a complete GA which remains invaluable for detecting unidentified problems, providing detailed guidance as to which interventions should be implemented and ultimately helping appropriate cancer treatment selection. Finally, previously reported results suggest potentially better diagnostic performance with the modified G8 than original G8 in identifying patients who need a complete GA, as indicated by results from the development study initially led in the ELCAPA cohort [122], but also confirmed in a recently published external validation study in which the modified G8 demonstrated good diagnostic performance and was found to outperform both gait speed and original G8 [75]. While these results need to be further confirmed in other large cohort studies, all those elements combined may support wider utilization of the most recent modified G8 instrument.

## 6.5 Conclusion

The present analysis identified both G8 and modified G8 as strong and consistent predictors of overall survival in a large population of older patients with cancer, regardless of the metastatic status or cancer site. These findings strengthen the clinical utility of those two screening tools in the geriatric oncology setting.



## CHAPTER 7

---

### **General Discussion and Perspectives**

The findings of this thesis and their corresponding discussions were reported in the previous four chapters. The following discussion chapter highlights elements not mentioned before relating to the **conceptual design** and **validation results** of other screening tools compared to the modified G8. The section also focuses on the **implications** of our findings for clinical practice, summarizes the general **strengths** and **limitations** and presents the **conclusions** and some **perspectives** for further research.

## 7.1 The modified G8 compared to other screening tools

### 7.1.1 Conceptual design

Several screening tools have been evaluated and validated in older patients with cancer for detection of impairments on GA [56]. However, some screening tools were initially developed for a different purpose or for populations other than older patients with cancer.

The Karnofsky Performance Status [58] and the ECOG Performance Status [59] were compared with GA in 2 and 1 study, respectively, despite important differences in their conceptual framework and objective. The Karnofsky index was introduced in 1949, at the beginning of cancer chemotherapy, with the purpose of allowing physicians to measure the impact of cancer and its treatment on function in some of the first patients with cancer to receive chemotherapy in the USA; in 1960, an alternate scale of

performance status was developed by the Eastern Cooperative Oncology Group, derived from the KPS to consistently assess the impact of a person's disease on their daily living abilities.

The first instrument that was actually designed as a frailty screening tool was the one designed by Fried and colleagues. They used data from the Cardiovascular Health Study (CHS), a prospective, observational study of men and women 65 years and older, to develop a screening tool to identify frail persons at high risk of adverse health-related outcomes in the clinical setting [22]. They formulated specific criteria to operationalize the definition of frailty, based on a conceptual framework [176].

The GFI is another screening instrument developed to determine a person's level of frailty in older people [62]. On the basis of literature and theory, 22 items were formulated and tested in a sample of 275 older people aged 65 and over. After internal consistency analysis and principal component analysis, 15 items were selected. Cut-off scores for frailty profiles were determined by geriatric experts.

In 2008, the Study of Osteoporotic Fractures (SOF) Research Group developed an index to measure pre-frailty and frailty status [84]. Based on the physiologic domains most frequently cited in the literature, findings from previous studies that evaluated the predictive validity of individual components and suitability of assessment of components in a busy clinical practice setting, authors proposed a short 3-item instrument.

While the three above instruments were developed to identify frail persons, frailty status or to determine a person's level of frailty, other instruments such as the Barber questionnaire, the fTRST, the ISAR and the VES-13 were designed for different purposes, namely risk for dependence or functional decline, death, repeat emergency department

(ED) visits and hospitalizations.

In the early 1980s, Barber and colleagues [76] developed a questionnaire aiming to identify elderly persons at risk for dependence. Nine questions from the GA were chosen after discussions with members of a University Department of Geriatric Medicine in England and several trial runs of the questionnaire to ensure that each question was unambiguous and easily understood and represented an important aspect of health.

The VES-13 [79] was developed to identify older people at risk for health deterioration. It is a risk prediction tool designed to predict the risk of death and functional decline in older community-dwelling people aged 65 years and older. It was derived through a methodologically robust process, whereby variables with potential predictive power were identified from the United States Medicare database and different models tested for relevant outcomes.

The original TRST screening tool was developed to identify community-dwelling older people at risk for subsequent return ED visits, unplanned hospitalizations, or nursing home placement [177]. Risk factors for the different outcomes were first reviewed in the literature. Then, they were reviewed for clinical applicability and feasibility in the ED setting by an expert panel of physicians, nurses and social workers, all specialized in gerontology. Five risk factors were chosen for the initial instrument, which was extended after a pilot study with an item concerning professional recommendations. The Flemish version was modified at the University Hospitals of Leuven, from the initial 5 item version [77]. It was used to identify older patients with a geriatric profile, who can benefit from an interdisciplinary intervention with extended assessment. The first exploratory study in 55 elderly cancer patients showed a better cut-off score of 1 instead of 2 for non-cancer patients.

The ISAR was developed in Canada in 1999 from a cohort of ED patients aged 65 and older to identify elderly patients at risk of adverse health outcomes during the 6 months after the ED visit [78]. Twenty-seven potential screening items were derived first from a literature review of risk factors for mortality, institutionalization, and functional decline, and then by expert panel discussions (multidisciplinary group of hospital and community-based health professionals). They were completed as part of the ED interview and their associations with each outcome was examined. The best subset of items was identified following a statistical procedure of variable selection.

The aCGA was the first screening tool developed to identify older cancer patients who might benefit from administration of a complete geriatric assessment [81]. It was based on a chart review of more than 500 patients aged 70 years and over, seen by the Senior Adult Oncology Program (SAOP) at the H. Lee Moffitt Cancer Center at the University of South Florida. The selection of items was based on psychometric criteria; those items within each scale that showed the highest item-to-total correlation were selected and a cut-off value was identified for each domain.

As the aCGA, 4 other screening tools were developed to help identifying older cancer patients who might benefit from a complete GA:

The OGS was developed in France in 2006 by the oncogeriatric team of the University Hospital of Poitiers [64], as a simple decision-making algorithm for the oncologists with a view to identify patients aged 75 years and over whose personalized treatment would be optimized by undergoing a GA. The tool had to be easy to apply so that its criteria could be incorporated into the questions asked at the oncological consultation.

The SAOP2 screening questionnaire was developed in 2008 by the multidisciplinary

clinical team of the SAOP at Moffitt to determine when a multidisciplinary evaluation by a geriatric oncology team was required in new cancer patients [83].

The Gerhematolim tool was developed in France in 2009 for older patients aged 70 years and older suffering from hematological malignancies [178]. It was made from SIOG and NCCN recommendations, the clinical experience of the multidisciplinary team of the haematologic network of Limousin and following French medical and economic specificities.

The G8 is the most recent screening tool specifically developed in an older population with cancer. The development study was based on a multicenter prospective cohort of 364 cancer patients aged 70 years and older from 12 French regional centers [61]. Following the multidisciplinary expertise of geriatricians and oncologists, specific items from the MNA were selected to constitute the core of the G8, plus an indicator of age. Considered elements for this selection were their expected correlations with most dimensions of the GA [60], and results from a preliminary analysis [179] which investigated factors associated with early death risk (death within 6 months of treatment initiation) in elderly patients with cancer under chemotherapy. Among GA data, MNA was the most predictive of early death in a logistic regression model. Cut-off values for a geriatric risk profile were determined by ROC analysis.

Several instruments were designed to be self-reported questionnaires (Barber, VES-13, ISAR, GFI), others to be used by the nursing personnel (TRST, Gerhematolim, G8), or a combination of these two (SAOP2), and others to be filled in by the clinician, i.e. the treating oncological specialist (Fried, aCGA, OGS, SOF, KPS, ECOG-PS). In any case, most instruments evoke a sense of simplicity, ease to use and quick tool. For example, the GFI was developed as a short and easy to use screening instrument; the SOF, as a simple

frailty index using only 3 components; the TRST was designed so it would only take a minute or two.

Although some instruments were designed following a specific statistical approach, most of them were empirically derived, either based on a conceptual framework like the frailty phenotype as described by Fried and colleagues, or following international or expert-based recommendations or defined by a multidisciplinary team (Barber questionnaire, OGS, SAOP2, Gerhematolim). For some tools, the selection of the items involved a two-step process of literature review of factors associated with the outcome of interest and expert panel consensus or authors discussions (TRST, SOF, G8). Only three tools combined literature review, expert panel and statistical methodology (ISAR, GFI and VES-13), but none developed specifically for an older cancer population. The aCGA for instance was only based on psychometric criteria and included items from only 4 domains. Further, the criteria used to define the frail person or what would be considered as an indicator of a geriatric risk profile varied widely between the tools and were mostly established by experts.

Compared to the development of other screening tools that have been used in older cancer patients for detection of impairments on GA, the modified G8 was developed following a step-by-step approach using literature, clinical expertise and a statistical process involving several methods, namely MCA, univariate and multivariate logistic regression, multiple imputation, final weights computation, internal and external validation. This systematic approach helped us to propose a hopefully relevant alternative to the original G8 aimed to identify patients in need of a complete GA, when the latter had been found to be the most robust screening tool in terms of sensitivity and prognostic value for outcome measures [56].

### 7.1.2 Predictive value on outcomes other than GA and mortality

Some of the tools mentioned above have been validated on specific outcomes [56], such as risk of falls (Fried criteria and SOF), recurrent fractures (Fried criteria and SOF), functional impairment (VES-13 and ISAR), functional decline / disability (Fried criteria, SOF, VES-13, fTRST, ISAR and G8), chemotherapy-related toxicity (GFI, VES-13 and G8) and overall survival (Fried criteria, GFI, SOF, VES-13, fTRST and G8). However, not all the populations concerned patients with cancer. Many validation studies have been performed in the general older population or in older patients presenting at ED. Others are based on populations with a specific cancer location and only few studies have been performed in older populations with various cancer sites [72,89,91,95,180], some based on very small sample sizes (i.e. 21 patients [91]). Therefore, the association of screening tools with these outcomes needs to be explored in larger prospective studies. The modified G8 has so far been studied in terms of overall survival, with promising results, comparable to those of the original G8 and the ECOG-PS. Its predictive value for other relevant outcomes should be further studied.

### 7.1.3 Direct comparisons between instruments

Direct comparisons between screening tools have rarely been conducted despite their potential interest. One study has recently evaluated the agreement between 35 frailty scores – including the G8 – in the general population and found a high level of heterogeneity between instruments to define frailty and, as a result, frequent and highly variable disagreements between tools [181]. In the older population with cancer and as previously discussed in the sections relating to development of the modified G8 and its prognostic value, there is a current need for direct comparisons between multiple

instruments based on common populations. Such comparisons could also benefit from more advanced analytic approaches based on regression modeling of tests' accuracy and adjustment for potentially influential covariates to provide more detailed information and balanced comparisons between tools [182]. Research projects that specifically aim at or incorporate sub-objectives of comparing instruments should be supported. Results from such studies will provide valuable insights on the advantages and drawbacks of screening tools and will help building better informed guidelines for real life setting.

## **7.2 Implications for practice**

This thesis has provided insights into how the optimized screening tool modified G8 could guide healthcare professionals on ways to improve the management of older patients with cancer. Our findings indicate that the modified G8 could effectively be used as part of several practical approaches that require further investigation. In this section, the possible implications are identified and discussed.

### **7.2.1 Two-step approach**

As mentioned before, although GA is considered the most appropriate way to examine the overall health status of the older patient with cancer, it remains a complex and interdisciplinary approach that requires considerable time to complete; in oncology settings, where 60%–70% of newly diagnosed cancer patients are older, it might not be feasible to give all older cancer patients a complete GA because of limited resources. A two-step screening process seems therefore appropriate: a first step where all patients are screened using the modified G8, previously shown to have good diagnostic

properties for identifying different health profiles evocative of frailty; and a second step in which those patients who scored 6 or above are seen for a complete GA. The SIOG has already recommended a similar approach for older patients with prostate cancer in which patients are first screened to determine their health status, followed by a simplified GA and a complete GA when necessary [50,51]. SIOG guidelines recommend a systematic use of the G8 screening instrument as a first step of the decision-making process. From our results, such process may have room for improvement, and the modified G8 may constitute a possible alternative. When both scores were compared to GA, defined as  $\geq 1$  impairment among the seven tests, in the ELCAPA and the ONCODAGE cohorts, the modified G8 questionnaire was found to have higher accuracy than the original G8 with higher AUROCs. In the Physical Frailty in Elderly Cancer patients (PF-EC) cohort [75], both tools had similar sensitivities when compared to GA ( $\geq 1$  impairment), but the modified G8 showed better specificity. When they were compared against the first classification of SIOG [49], distinguishing fit from non-fit patients, in the ELCAPA cohort, the modified G8 score showed better sensitivity, specificity and AUROC than the original G8, encouraging the actual use of this new tool in a two-step approach.

However and regardless of the screening tool considered, the effectiveness of such an approach has not yet been established and needs to be validated in randomized controlled trials. The current clinical trial PREPARE [183] has been designed as such. First, all older cancer patients treated at the participating centers are screened with the G8 instrument. Then, those patients considered in need of GA (altered score:  $\leq 14$ ) are included in the main trial, which primary objective is to assess the efficacy of geriatric intervention in the management of older patients with cancer, compared to usual care, using a co-primary main endpoint encompassing 1-year overall survival and health related quality of life, providing additionally a description of vital status of patients with

a normal G8-score. In this context, the modified G8 could be used as well, since it has shown already good performances for discriminating fit from unfit patients with better diagnostic properties compared to the original G8.

### 7.2.2 Modified G8 as part of a 'Geriatric Mini Data-Set'

The lack of standardization in classifying patients based on GA findings (i.e. fit or frail) has led to difficulties regarding cross-study comparisons. For instance, the prevalence of what is considered an abnormal GA in any population, correlates positively with the number of conditions evaluated and strongly depends on the selected tools and cutoffs for defining impairment. Despite a general agreement concerning the domains that a GA should comprise, there are several different tools used to evaluate these domains [28]. Therefore, some researches have focused on the standardization of these assessment tools in the context of clinical trials [184,185]. Indeed, older patients are often poorly represented in clinical trials in oncology [186-188]. Thus, incorporating minimal data collection to be included in any clinical trial involving older people seems vital in order to achieve harmonization and overcome the little evidence from insufficient clinical research in this population. A *Geriatric Minimum Data Set* (GMDS) [184] and a *Mini data-set* for older patients with cancer [185] have been proposed as minimum sets of information to allow a standardized description of the older person participating in clinical research. The GMDS, developed by a multidisciplinary task force group of investigators from different leading Research Centers in Europe, is a 25-item data set covering important health domains, and proposes to include a frailty index. However, this data set remains impractical and time-consuming, given the multiple evaluations to be performed. The French Mini data-set was developed from another perspective, providing appropriate measures with only 9 items that would greatly

facilitate patient assessment for clinical research. In this sense, simple indicators of the patient health status, such as items from the original and modified G8, could be considered to be included as part of this minimum set of information, enhancing opportunity for research in populations of older patients with cancer. The utility of this mini data-set incorporating the screening tool might also be evaluated in terms of routine management in care facilities. The amount of time necessary to collect data is likely to be within an acceptable range thanks to the shortness of the tool, yet the feasibility of implementing the instrument would require further evaluation.

### 7.2.3 Modified G8 as a prognostic tool

Several studies have demonstrated independent associations linking multidimensional impairments explored by a complete GA and overall survival in older patients with cancer (i.e. functional, nutritional, comorbidities, cognitive and psychosocial domains) [28,55]. Thus, a cancer specific *GA-based prognostic instrument* for one-year survival was developed to help guide treatment plan in older cancer patients [189] and it demonstrated to be a good predictive tool, with a very good discriminatory power (C-statistic of 0.87). However, the tool needs to be externally validated to confirm its predictive value.

On the other hand, focus on screening instruments has recently shifted from predicting impairments to predicting endpoints relevant to older patients with cancer, such as overall survival [56]. The modified G8 showed to be strongly associated with an increased mortality risk within the first and three years after inclusion, which remained significant after multivariable adjustments for age, gender, metastatic status, anti-cancer treatment and type of malignancy. Moreover, this new tool takes into account 6 GA

domains related to functional, cognitive, nutritional and self-rated health status, comorbidities, and polypharmacy, each evaluated with one simple item. Although it was not developed for this purpose, it showed a very good prognostic performance for overall survival. This information may be especially helpful in clinical decision making in this population. It may therefore be appropriate to classify patients enrolled in clinical trials, selecting more homogeneous prognostic groups of patients, since our results demonstrated that classifying patients in classes of increasing risk had a graded relationship with survival. A possible strategy could be to recommend only palliative treatment to those older cancer patients with a very high modified G8 score (i.e. >30), considered to have very poor survival and at least one GA impairment. In the same way, a standard therapy could be recommended to those patients with a low modified G8 score (i.e. <6). Concerning the intermediated groups, a specific adapted therapy may be recommended. However, to determine whether geriatric interventions are needed in order to better tolerate therapy, a GA seems essential.

Since the modified G8 demonstrated to be an important prognostic tool for survival in older patients with cancer, its integration at the time of treatment decisions could be beneficial. Its utility as a prognostic tool for helping physicians to decide treatment options must be validated in the context of clinical trials.

### **7.3 Strengths and limitations**

This thesis underlines the interest of applying complementary approaches for the development of a new screening tool, accounting both for statistical and clinical relevance concerns in a view to optimize the usability and actual contribution of the instrument to clinical practice. Our findings will hopefully provide valuable insights for

the clinician with respect to the pragmatic implementation and understanding of frailty in older patients with cancer. Thus, one of the strengths of this work relates to the methodological approach used to develop and validate the modified G8 screening tool, using recommended guidelines and appropriate statistical procedures. We used a well-defined reference standard to determine patient's health status, and comparable to that defined in the first evaluation study of the original G8 tool. Further, data were based on two large oncogeriatric prospective cohort studies including various solid and hematological cancers. For the validation study, we chose a national multicenter cohort including only first-line cancer treatment patients from a large number of investigating centers, including community hospitals, and seemed to be representative of older patients with newly diagnosed cancer. Finally, multiple sensitivity analyses were performed regarding its predictive value for survival, as well as the various reference definitions tested according to different approaches in order to test the robustness of the instrument in terms of diagnostic properties. Additionally, results were systematically compared with the original G8, and were performed globally and by cancer site.

Among the general limitations of the present work, it should be noticed that for each study we excluded patients with no complete data on either the reference definition tested or the screening tool, which could have led to selection biased in some studies where we found significant differences between included and excluded patients concerning demographic and clinical characteristics. Another limitation was the non-available detail on the modalities of specific variables that could potentially have been included in the new tool, such as ADL or IADL modalities or modalities from other validated scales. Finally, we could not compare our results directly to other pertinent screening tools as they were unavailable in the data set.

## CONCLUSIONS AND PERSPECTIVES FOR FURTHER RESEARCH

---

Management of older patients with cancer represents a significant challenge for clinical care, due to the wide heterogeneity and complexity of the health issues encountered in this population. It is therefore important to identify those frail patients who are at risk of poor clinical outcomes, and in whom interventions or treatment modifications are needed.

Based on the findings of this work, the proposed modified G8 may represent a credible candidate among available screening tools for helping the physician to identify frail older patients with cancer requiring a complete GA. Our results have been confirmed in two independent populations. Moreover, the modified G8 seems to be an appropriate tool to identify several profiles suggesting frailty, regardless of the definition which has been continuously debated over the past decades. Since the choice of the most appropriate definition to operationalize this concept remains difficult and without consensus, we hope our findings may have the potential to offer a practical response for daily practice with an instrument able to detect any potential risk problem regardless of the definition. In this context, the modified G8 demonstrated good diagnostic performances when tested against six different reference definitions. Furthermore, the modified G8 predicted overall survival strongly in a wide variety of tumor types, irrespective of treatment choice and metastatic status.

All these findings support the potential value of the optimized 6-item version, the modified G8, to guide the oncologist in the determination of the best management strategy for these patients.

Considering that our work is the first to report on the modified G8, further studies will be necessary to corroborate our findings. In particular, it would be of special interest to implement the explicit modified G8 in other independent and larger populations allowing us to confirm its diagnostic and prognostic properties regarding outcomes others than overall survival, namely functional decline, risk of falls, treatment-related toxicity, completion of therapy and quality of life. The reproducibility of the instrument should additionally be addressed in a future study. Other steps will include the assessment of effectiveness of repeated screening during the treatment process, and determine if geriatric management integrating the modified G8 ultimately improves patient health outcomes.

## REFERENCES

---

- [1] Institut national de la statistique et des études économiques (Insee). Bilan démographique 2016. <https://www.insee.fr/fr/statistiques/2554860>. Accessed June 30, 2017.
- [2] Institut national du cancer. © *Les cancers en France, édition 2016*. Collection Les Données, avril 2017.
- [3] Berger NA, Savvides P, Koroukian SM, et al. Cancer in the elderly. *Transactions of the American Clinical and Climatological Association*. 2006;117:147-155; discussion 155-146.
- [4] United States Census Bureau's International Data Base. <https://www.census.gov/population/international/data/idb/worldpop.php>. Accessed June 22 2017.
- [5] United States National Institute on Aging, World Health Organization. *Global Health and Aging 2011*. NIH Publication no. 11-7737.
- [6] United Nations, Department of Economic and Social Affairs, Population Division. *World Population Ageing 2015 (ST/ESA/SER.A/390)*. 2015.
- [7] Institut national du cancer. *Les cancers en France en 2016, l'essentiel des faits et chiffres*. Collection Les données, 2017.
- [8] Howlader N NA, Krapcho M, Garshell J, Miller D, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). *SEER Cancer Statistics Review, 1975-2012*. National Cancer Institute. Bethesda, MD, [http://seer.cancer.gov/csr/1975\\_2012/](http://seer.cancer.gov/csr/1975_2012/); based on November 2014 SEER data submission, posted to the SEER web site, April 2015.
- [9] Howlader N, Noone A, Krapcho M, et al. *SEER Cancer Statistics Review, 1975-2011*. Bethesda, MD.
- [10] Statistics CCSsACoC. *Canadian Cancer Statistics 2014*. Toronto, ON: Canadian Cancer Society, 2014.
- [11] Viganò A, Morais JA. The elderly patient with cancer: A holistic view. *Nutrition*. 2015;31(4):587-589.
- [12] Baijal P, Periyakoil V. Understanding Frailty in Cancer Patients. *The Cancer Journal*. 2014;20(5):358-366.
- [13] Wedding U, Honecker F, Bokemeyer C, Pientka L, Hoffken K. Tolerance to chemotherapy in elderly patients with cancer. *Cancer control : journal of the Moffitt Cancer Center*. 2007;14(1):44-56.
- [14] Vincent H, Caillet P, Culine S, Paillaud E. Spécificités et recommandations concernant la chimiothérapie chez le sujet âgé : toxicité rénale et cardiaque. *NPG Neurologie - Psychiatrie - Gériatrie*. 2011;11(66):250-258.
- [15] Pal SK, Katheria V, Hurria A. Evaluating the Older Patient with Cancer: Understanding Frailty and the Geriatric Assessment. *Ca-Cancer J Clin*. 2010;60(2):120-132.
- [16] Clegg A, Young J, Iliffe S, Rikkert MO, Rockwood K. Frailty in elderly people. *Lancet*. 2013;381(9868):752-762.
- [17] Walston J, Hadley EC, Ferrucci L, et al. Research agenda for frailty in older adults: toward a better understanding of physiology and etiology: summary from the American Geriatrics Society/National Institute on Aging Research Conference on Frailty in Older Adults. *Journal of the American Geriatrics Society*. 2006;54(6):991-1001.

- [18] Morley JE, Vellas B, van Kan GA, et al. Frailty consensus: a call to action. *Journal of the American Medical Directors Association*. 2013;14(6):392-397.
- [19] Kelaiditi E, Cesari M, Canevelli M, et al. Cognitive frailty: rational and definition from an (I.A.N.A./I.A.G.G.) international consensus group. *The journal of nutrition, health & aging*. 2013;17(9):726-734.
- [20] Gobbens RJ, Luijkx KG, Wijnen-Sponselee MT, Schols JM. In search of an integral conceptual definition of frailty: opinions of experts. *Journal of the American Medical Directors Association*. 2010;11(5):338-343.
- [21] Hurria A, Dale W, Mooney M, et al. Designing therapeutic clinical trials for older and frail adults with cancer: U13 conference recommendations. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2014;32(24):2587-2594.
- [22] Fried LP, Tangen CM, Walston J, et al. Frailty in older adults: evidence for a phenotype. *J Gerontol A Biol Sci Med Sci*. 2001;56(3):M146-156.
- [23] Rockwood K, Song X, MacKnight C, et al. A global clinical measure of fitness and frailty in elderly people. *Canadian Medical Association Journal*. 2005;173(5):489-495.
- [24] Blodgett J, Theou O, Kirkland S, Andreou P, Rockwood K. Frailty in NHANES: Comparing the frailty index and phenotype. *Archives of Gerontology and Geriatrics*. 2015;60(3):464-470.
- [25] Theou O, Brothers TD, Mitnitski A, Rockwood K. Operationalization of Frailty Using Eight Commonly Used Scales and Comparison of Their Ability to Predict All-Cause Mortality. *Journal of the American Geriatrics Society*. 2013;61(9):1537-1551.
- [26] Rubenstein LZ, Stuck AE, Siu AL, Wieland D. Impacts of geriatric evaluation and management programs on defined outcomes: overview of the evidence. *Journal of the American Geriatrics Society*. 1991;39(9 Pt 2):8S-16S; discussion 17S-18S.
- [27] Extermann M, Aapro M, Bernabei R, et al. Use of comprehensive geriatric assessment in older cancer patients: recommendations from the task force on CGA of the International Society of Geriatric Oncology (SIOG). *Crit Rev Oncol Hematol*. 2005;55(3):241-252.
- [28] Wildiers H, Heeren P, Puts M, et al. International Society of Geriatric Oncology Consensus on Geriatric Assessment in Older Patients With Cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2014;32(24):2595-2603.
- [29] National Comprehensive Cancer Network. Older Adult Oncology (Version 2.2017 Updates). [https://www.nccn.org/professionals/physician\\_gls/pdf/senior.pdf](https://www.nccn.org/professionals/physician_gls/pdf/senior.pdf). Accessed June 19, 2017.
- [30] Extermann M, Meyer J, McGinnis M, et al. A comprehensive geriatric intervention detects multiple problems in older breast cancer patients. *Crit Rev Oncol Hematol*. 2004;49(1):69-75.
- [31] Corre R, Greillier L, Caër HL, et al. Use of a Comprehensive Geriatric Assessment for the Management of Elderly Patients With Advanced Non-Small-Cell Lung Cancer: The Phase III Randomized ESOGIA-GFPC-GECP 08-02 Study. *Journal of Clinical Oncology*. 2016;34(13):1476-1483.
- [32] Ferrat E, Paillaud E, Laurent M, et al. Predictors of 1-Year Mortality in a Prospective Cohort of Elderly Patients With Cancer. *J Gerontol A Biol Sci Med Sci*. 2015;70(9):1148-1155.
- [33] Laurent M, Paillaud E, Tournigand C, et al. Assessment of solid cancer treatment feasibility in older patients: a prospective cohort study. *The oncologist*. 2014;19(3):275-282.

- [34] Clough-Gorr KM, Stuck AE, Thwin SS, Silliman RA. Older Breast Cancer Survivors: Geriatric Assessment Domains Are Associated With Poor Tolerance of Treatment Adverse Effects and Predict Mortality Over 7 Years of Follow-Up. *Journal of Clinical Oncology*. 2010;28(3):380-386.
- [35] Caillet P, Canoui-Poitrine F, Vouriot J, et al. Comprehensive Geriatric Assessment in the Decision-Making Process in Elderly Patients With Cancer: ELCAPA Study. *Journal of Clinical Oncology*. 2011;29(27):3636-3642.
- [36] Kalsi T, Babic-Illman G, Ross PJ, et al. The impact of comprehensive geriatric assessment interventions on tolerance to chemotherapy in older people. *British journal of cancer*. 2015;112(9):1435-1444.
- [37] Chaibi P, Magne N, Breton S, et al. Influence of geriatric consultation with comprehensive geriatric assessment on final therapeutic decision in elderly cancer patients. *Crit Rev Oncol Hemat*. 2011;79(3):302-307.
- [38] Puts MT, Hardt J, Monette J, Girre V, Springall E, Alibhai SM. Use of geriatric assessment for older adults in the oncology setting: a systematic review. *Journal of the National Cancer Institute*. 2012;104(15):1133-1163.
- [39] Mohile SG, Velarde C, Hurria A, et al. Geriatric Assessment-Guided Care Processes for Older Adults: A Delphi Consensus of Geriatric Oncology Experts. *Journal of the National Comprehensive Cancer Network : JNCCN*. 2015;13(9):1120-1130.
- [40] Stuck AE, Siu AL, Wieland GD, Adams J, Rubenstein LZ. Comprehensive geriatric assessment: a meta-analysis of controlled trials. *Lancet*. 1993;342(8878):1032-1036.
- [41] Ellis G, Whitehead MA, Robinson D, O'Neill D, Langhorne P. Comprehensive geriatric assessment for older adults admitted to hospital: meta-analysis of randomised controlled trials. *Bmj*. 2011;343:d6553.
- [42] McCorkle R, Strumpf NE, Nuamah IF, et al. A specialized home care intervention improves survival among older post-surgical cancer patients. *Journal of the American Geriatrics Society*. 2000;48(12):1707-1713.
- [43] Goodwin JS, Satish S, Anderson ET, Nattinger AB, Freeman JL. Effect of nurse case management on the treatment of older women with breast cancer. *Journal of the American Geriatrics Society*. 2003;51(9):1252-1259.
- [44] Rao AV, Hsieh F, Feussner JR, Cohen HJ. Geriatric evaluation and management units in the care of the frail elderly cancer patient. *J Gerontol A Biol Sci Med Sci*. 2005;60(6):798-803.
- [45] Horgan AM, Leighl NB, Coate L, et al. Impact and feasibility of a comprehensive geriatric assessment in the oncology setting: a pilot study. *American journal of clinical oncology*. 2012;35(4):322-328.
- [46] Balducci L, Extermann M. Management of the frail person with advanced cancer. *Crit Rev Oncol Hematol*. 2000;33(2):143-148.
- [47] Winograd CH, Gerety MB, Chung M, Goldstein MK, Dominguez F, Jr., Vallone R. Screening for frailty: criteria and predictors of outcomes. *Journal of the American Geriatrics Society*. 1991;39(8):778-784.
- [48] Balducci L, Extermann M. Management of cancer in the older person: a practical approach. *The oncologist*. 2000;5(3):224-237.
- [49] Droz JP, Balducci L, Bolla M, et al. Management of prostate cancer in older men: recommendations of a working group of the International Society of Geriatric Oncology. *BJU international*. 2010;106(4):462-469.
- [50] Droz JP, Aapro M, Balducci L, et al. Management of prostate cancer in older patients: updated recommendations of a working group of the International Society of Geriatric Oncology. *The Lancet. Oncology*. 2014;15(9):e404-414.

- [51] Droz JP, Albrand G, Gillessen S, et al. Management of Prostate Cancer in Elderly Patients: Recommendations of a Task Force of the International Society of Geriatric Oncology. *European urology*. 2017.
- [52] Yokom DW, Kenis C, Alibhai SM, Puts M. Screening Tools in Geriatric Oncology. In: Korc-Grodzicki B, Tew WP, eds. *Handbook of Geriatric Oncology, Practical Guide to Caring for the Older Cancer Patient*. New York: Demos Medical Publishing; 2017:119-130.
- [53] Wilson JMG, Jungner G. *Principles and practice of screening for disease*. Geneva,: World Health Organization; 1968.
- [54] World Health Organization. Screening for various cancers. <http://who.int/cancer/detection/variouscancer/en/>. Accessed 8 June, 2017.
- [55] Caillet P, Laurent M, Bastuji-Garin S, et al. Optimal management of elderly cancer patients: usefulness of the Comprehensive Geriatric Assessment. *Clinical interventions in aging*. 2014;9:1645-1660.
- [56] Decoster L, Van Puyvelde K, Mohile S, et al. Screening tools for multidimensional health problems warranting a geriatric assessment in older cancer patients: an update on SIOG recommendations. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2014;26(2):288-300.
- [57] Hamaker ME, Jonker JM, de Rooij SE, Vos AG, Smorenburg CH, van Munster BC. Frailty screening methods for predicting outcome of a comprehensive geriatric assessment in elderly patients with cancer: a systematic review. *Lancet Oncology*. 2012;13(10):E437-E444.
- [58] Karnofsky DA, Burchenal J. H. The Clinical Evaluation of Chemotherapeutic Agents in Cancer *Evaluation of Chemotherapeutic Agents*. C. M. MacLeod ed. New York: Columbia University Press; 1949:191-205.
- [59] Oken MM, Creech RH, Tormey DC, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. *American journal of clinical oncology*. 1982;5(6):649-655.
- [60] Soubeyran P, Bellera CA, Gregoire F, et al. Validation of a screening test for elderly patients in oncology. *Journal of Clinical Oncology*. 2008;26(15\_suppl):20568-20568.
- [61] Bellera CA, Rainfray M, Mathoulin-Pelissier S, et al. Screening older cancer patients: first evaluation of the G-8 geriatric screening tool. *Annals of Oncology*. 2012;23(8):2166-2172.
- [62] Steverink N, Slaets J, Schuurmans H, Van Lis M. Measuring frailty. Development and testing of the Groningen Frailty Indicator (GFI). *Gerontologist*. 2001;41(1):236.
- [63] Schuurmans H, Steverink N, Lindenberg S, Frieswijk N, Slaets JP. Old or frail: what tells us more? *J Gerontol A Biol Sci Med Sci*. 2004;59(9):M962-965.
- [64] Valero S, Migeot V, Bouche G, et al. Who needs a comprehensive geriatric assessment? A French Onco-Geriatric Screening tool (OGS). *J Geriatr Oncol*. 2011;2(2):130-136.
- [65] Fargeas JB, Vigneras B, Marin B, Bosselut M, Bouzogne C, Bordessoule D. Reproducibility of the screening tool GERHEMATOLIM in geriatric patients over 70 years with malignant haemopathy. *Crit Rev Oncol Hemat*. 2009;72:S24-S24.
- [66] Fargeas JB. *Validation d'un test de dépistage gériatrique qbrégé adapté à des patients de plus de 70 ans atteints d'hémopathie maligne* [Thèse de médecine]. Limoges: Faculté de Médecine, Université de Limoges; 2009.
- [67] Hamaker ME, Mitrovic M, Stauder R. The G8 screening tool detects relevant geriatric impairments and predicts survival in elderly patients with a haematological malignancy. *Annals of hematology*. 2014;93(6):1031-1040.

- [68] Smets IH, Kempen GI, Janssen-Heijnen ML, Deckx L, Buntinx FJ, van den Akker M. Four screening instruments for frailty in older patients with and without cancer: a diagnostic study. *BMC geriatrics*. 2014;14:26.
- [69] Kenig J, Richter P, Zychiewicz B, Olszewska U. Vulnerable Elderly Survey 13 as a screening method for frailty in Polish elderly surgical patient--prospective study. *Polski przegląd chirurgiczny*. 2014;86(3):126-131.
- [70] Holmes HM, Des Bordes JK, Kebriaei P, et al. Optimal screening for geriatric assessment in older allogeneic hematopoietic cell transplantation candidates. *J Geriatr Oncol*. 2014;5(4):422-430.
- [71] Velghe A, Petrovic M, De Buyser S, Demuynck R, Noens L. Validation of the G8 screening tool in older patients with aggressive haematological malignancies. *European journal of oncology nursing : the official journal of European Oncology Nursing Society*. 2014;18(6):645-648.
- [72] Augschoell J, Kemmler G, Hamaker ME, Stauder R. PPT and VES-13 in elderly patients with cancer: evaluation in multidimensional geriatric assessment and prediction of survival. *J Geriatr Oncol*. 2014;5(4):415-421.
- [73] Kenig J, Zychiewicz B, Olszewska U, Richter P. Screening for frailty among older patients with cancer that qualify for abdominal surgery. *J Geriatr Oncol*. 2015;6(1):52-59.
- [74] Hentschel L, Rentsch A, Lenz F, et al. A Questionnaire Study to Assess the Value of the Vulnerable Elders Survey, G8, and Predictors of Toxicity as Screening Tools for Frailty and Toxicity in Geriatric Cancer Patients. *Oncology research and treatment*. 2016;39(4):210-216.
- [75] Pamoukdjian F, Canoui-Poitrine F, Longelin-Lombard C, et al. Diagnostic performance of gait speed, G8 and G8 modified indices to screen for vulnerability in older cancer patients: the prospective PF-EC cohort study. *Oncotarget*. 2017.
- [76] Barber JH, Wallis JB, McKeating E. A postal screening questionnaire in preventive geriatric care. *The Journal of the Royal College of General Practitioners*. 1980;30(210):49-51.
- [77] Kenis C, Geeraerts A, Braesl T, Milisen K, Flamaing J. The Flemish version of the Triage Risk Screening Tool (TRST): A multidimensional short screening tool for the assessment of elderly patients. *Crit Rev Oncol Hemat*. 2006;60:S31-S31.
- [78] McCusker J, Bellavance F, Cardin S, Trepanier S, Verdon J, Ardman O. Detection of Older People at Increased Risk of Adverse Health Outcomes After an Emergency Visit: The ISAR Screening Tool. *Journal of the American Geriatrics Society*. 1999;47(10):1229-1237.
- [79] Saliba D, Elliott M, Rubenstein LZ, et al. The Vulnerable Elders Survey: a tool for identifying vulnerable older people in the community. *Journal of the American Geriatrics Society*. 2001;49(12):1691-1699.
- [80] Overcash JA, Beckstead J, Moody L, Extermann M, Cobb S. The abbreviated comprehensive geriatric assessment (aCGA) for use in the older cancer patient as a prescreen: scoring and interpretation. *Crit Rev Oncol Hematol*. 2006;59(3):205-210.
- [81] Overcash JA, Beckstead J, Extermann M, Cobb S. The abbreviated comprehensive geriatric assessment (aCGA): a retrospective analysis. *Crit Rev Oncol Hemat*. 2005;54(2):129-136.
- [82] Extermann M. Senior Adult Oncology Program Screening Questionnaire SAOP2 <https://moffitt.org/for-healthcare-providers/clinical-programs-and-services/senior-adult-oncology-program/senior-adult-oncology-program-tools/>. Accessed 31/05/2017.

- [83] Extermann M GT, Tiffenberg G, Rich CJ. Validation of the Senior Adult Oncology Program (SAOP) 2 screening questionnaire. *Critical Reviews in Oncology / Hematology*. 2009;69(2):185 (abstr P124a).
- [84] Ensrud KE, Ewing SK, Taylor BC, et al. Comparison of 2 frailty indexes for prediction of falls, disability, fractures, and death in older women. *Archives of internal medicine*. 2008;168(4):382-389.
- [85] Soubeyran P, Bellera C, Goyard J, et al. Screening for Vulnerability in Older Cancer Patients: The ONCODAGE Prospective Multicenter Cohort Study. *PloS one*. 2014;9(12):e115060.
- [86] Pottel L, Boterberg T, Pottel H, et al. Determination of an adequate screening tool for identification of vulnerable elderly head and neck cancer patients treated with radio(chemo)therapy. *J Geriatr Oncol*. 2012;3(1):24-32.
- [87] Petit-Moneger A, Rainfray M, Soubeyran P, Bellera CA, Mathoulin-Pelissier S. Detection of frailty in elderly cancer patients: Improvement of the G8 screening test. *J Geriatr Oncol*. 2016;7(2):99-107.
- [88] Kenis C, Decoster L, Van Puyvelde K, et al. Performance of two geriatric screening tools in older patients with cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2014;32(1):19-26.
- [89] Stokoe JM, Pearce J, Sinha R, Ring A. G8 and VES-13 scores predict chemotherapy toxicity in older patients with cancer. *J Geriatr Oncol*. 2012;3:S81.
- [90] Luciani A, Biganzoli L, Colloca G, et al. Estimating the risk of chemotherapy toxicity in older patients with cancer: The role of the Vulnerable Elders Survey-13 (VES-13). *J Geriatr Oncol*. 2015;6(4):272-279.
- [91] van Fraeyenhove F, Baitar A, De Vos M, et al. Prediction of chemotherapy toxicity by the Groningen Frailty Index (GFI) and the Comprehensive Geriatric Assessment (CGA) in Elderly cancer patients (pts): an interim Analysis. *Annals of Oncology*; 2010.
- [92] Liuu E, Canoui-Poitrine F, Tournigand C, et al. External validation of the G-8 geriatric screening tool to identify vulnerable elderly cancer patients: The ELCAPA-02 study. *J Geriatr Oncol*. 2012;3:S45-S46.
- [93] Kitamura H, Nagashima F, Miyajima K, et al. Continuous comprehensive geriatric assessment could predict the prognosis in elderly cancer patients. *J Geriatr Oncol*. 2013;4:S82-S83.
- [94] Aaldriks AA, Giltay EJ, le Cessie S, et al. Prognostic value of geriatric assessment in older patients with advanced breast cancer receiving chemotherapy. *Breast*. 2013;22(5):753-760.
- [95] Aaldriks AA, Maartense E, le Cessie S, et al. Predictive value of geriatric assessment for patients older than 70 years, treated with chemotherapy. *Crit Rev Oncol Hematol*. 2011;79(2):205-212.
- [96] Kristjansson SR, Rønning B, Hurria A, et al. A comparison of two pre-operative frailty measures in older surgical cancer patients. *J Geriatr Oncol*. 2012;3(1):1-7.
- [97] Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of Illness in the Aged. The Index of Adl: A Standardized Measure of Biological and Psychosocial Function. *Jama*. 1963;185:914-919.
- [98] Lawton MP, Brody EM. Assessment of older people: self-maintaining and instrumental activities of daily living. *Gerontologist*. 1969;9(3):179-186.
- [99] Mathias S, Nayak US, Isaacs B. Balance in elderly patients: the "get-up and go" test. *Archives of physical medicine and rehabilitation*. 1986;67(6):387-389.

- [100] Guigoz Y, Vellas B, PJ. G. Mini Nutritional Assessment: a practical assessment tool for grading the nutritional state of elderly patients. *Facts Res Gerontol* 1994;4(Suppl 2):15-59.
- [101] Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research*. 1975;12(3):189-198.
- [102] Clement JP, Nassif RF, Leger JM, Marchan F. [Development and contribution to the validation of a brief French version of the Yesavage Geriatric Depression Scale]. *L'Encephale*. 1997;23(2):91-99.
- [103] American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 4th ed. Washington, DC: American Psychiatric Association; 1994.
- [104] Sheikh JI, Yesavage JA. Geriatric Depression Scale (GDS): Recent evidence and development of a shorter version. . *Clinical Gerontologist*. 1986;5:165-172.
- [105] Yesavage JA. Geriatric Depression Scale. *Psychopharmacology bulletin*. 1988;24(4):709-711.
- [106] Miller MD, Towers A. *A Manual of Guidelines for Scoring the Cumulative Illness Rating Scale for Geriatrics (CIRS-G)*. Pittsburgh: PA, University of Pittsburgh; 1991.
- [107] Linn BS, Linn MW, Gurel L. Cumulative illness rating scale. *Journal of the American Geriatrics Society*. 1968;16(5):622-626.
- [108] Puts MT, Santos B, Hardt J, et al. An update on a systematic review of the use of geriatric assessment for older adults in oncology. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2014;25(2):307-315.
- [109] Extermann M, Hurria A. Comprehensive geriatric assessment for older patients with cancer. *Journal of Clinical Oncology*. 2007;25(14):1824-1831.
- [110] Hendriksen JM, Geersing GJ, Moons KG, de Groot JA. Diagnostic and prognostic prediction models. *Journal of thrombosis and haemostasis : JTH*. 2013;11 Suppl 1:129-141.
- [111] FIRTH D. Bias reduction of maximum likelihood estimates. *Biometrika*. 1993;80(1):27-38.
- [112] Lebart L, Monneau A, Warwick K. *Multivariate descriptive statistical analysis: correspondence analysis and related techniques*. New York 1984.
- [113] Cole TJ. Scaling and rounding regression coefficients to integers. *Appl Stat*. 1993;42:261-268.
- [114] Steyerberg E. *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. New York 2008.
- [115] J. A. Diagnostic tests: Moving beyond sensitivity and specificity: using likelihood ratios to help interpret diagnostic tests. *Australian Prescriber*. 2003;26(5):111-113.
- [116] Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York 1987.
- [117] Kristjansson SR, Nesbakken A, Jordhoy MS, et al. Comprehensive geriatric assessment can predict complications in elderly patients after elective surgery for colorectal cancer: a prospective observational cohort study. *Crit Rev Oncol Hematol*. 2010;76(3):208-217.
- [118] Soubeyran P, Fonck M, Blanc-Bisson C, et al. Predictors of early death risk in older patients treated with first-line chemotherapy for cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2012;30(15):1829-1834.
- [119] Petit-Monegera A, Rainfray M, Soubeyran P, Bellera C, Mathoulin-Pelissier S. Amélioration des performances d'un test de dépistage gériatrique à huit questions visant à détecter les patients âgés fragiles en cancérologie. *Rev Epidemiol Sante Publique*. 2014;62(S4):S118.

- [120] Baitar A, Van Fraeyenhove F, Vandebroek A, et al. Evaluation of the Groningen Frailty Indicator and the G8 questionnaire as screening tools for frailty in older patients with cancer. *J Geriatr Oncol.* 2013;4(1):32-38.
- [121] Liuu E, Canoui-Poitrine F, Tournigand C, et al. Accuracy of the G-8 geriatric-oncology screening tool for identifying vulnerable elderly patients with cancer according to tumour site: the ELCAPA-02 study. *J Geriatr Oncol.* 2014;5(1):11-19.
- [122] Martinez-Tapia C, Canoui-Poitrine F, Bastuji-Garin S, et al. Optimizing the G8 Screening Tool for Older Patients With Cancer: Diagnostic Performance and Validation of a Six-Item Version. *The oncologist.* 2016;21(2):188-195.
- [123] Moons KG, Kengne AP, Grobbee DE, et al. Risk prediction models: II. External validation, model updating, and impact assessment. *Heart.* 2012;98(9):691-698.
- [124] Collins GS, de Groot JA, Dutton S, et al. External validation of multivariable prediction models: a systematic review of methodological conduct and reporting. *BMC medical research methodology.* 2014;14:40.
- [125] Toll DB, Janssen KJ, Vergouwe Y, Moons KG. Validation, updating and impact of clinical prediction rules: a review. *Journal of clinical epidemiology.* 2008;61(11):1085-1094.
- [126] Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. *Journal of clinical epidemiology.* 2008;61(1):76-86.
- [127] Steyerberg E. *Clinical Prediction Models.* New York: Springer; 2009.
- [128] Van Hoorde K, Vergouwe Y, Timmerman D, Van Huffel S, Steyerberg EW, Van Calster B. Simple dichotomous updating methods improved the validity of polytomous prediction models. *Journal of clinical epidemiology.* 2013;66(10):1158-1165.
- [129] van Houwelingen HC. Validation, calibration, revision and combination of prognostic survival models. *Statistics in medicine.* 2000;19(24):3401-3415.
- [130] Debray TP, Vergouwe Y, Koffijberg H, Nieboer D, Steyerberg EW, Moons KG. A new framework to enhance the interpretation of external validation studies of clinical prediction models. *Journal of clinical epidemiology.* 2014.
- [131] Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology.* 2010;21(1):128-138.
- [132] Stukel TA. Generalized Logistic Models. *Journal of the American Statistical Association.* 1988;83(402):426-431.
- [133] Fraccaro P, van der Veer S, Brown B, et al. An external validation of models to predict the onset of chronic kidney disease using population-based electronic health records from Salford, UK. *BMC medicine.* 2016;14:104.
- [134] Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology.* 1982;143(1):29-36.
- [135] Gerds TA, Cai T, Schumacher M. The performance of risk prediction models. *Biometrical journal. Biometrische Zeitschrift.* 2008;50(4):457-479.
- [136] Marcin JP, Romano PS. Size matters to a model's fit. *Critical care medicine.* 2007;35(9):2212-2213.
- [137] Bertolini G, D'Amico R, Nardi D, Tinazzi A, Apolone G. One model, several results: the paradox of the Hosmer-Lemeshow goodness-of-fit test for the logistic regression model. *Journal of epidemiology and biostatistics.* 2000;5(4):251-253.
- [138] Paul P, Pennell ML, Lemeshow S. Standardizing the power of the Hosmer-Lemeshow goodness of fit test in large data sets. *Statistics in medicine.* 2013;32(1):67-80.
- [139] Allison P. Measures of fit for logistic regression. Paper 1485-2014. SAS Global Forum; 2014; Washington, DC.

- [140] Vergouwe Y, Steyerberg EW, Eijkemans MJC, Habbema JDF. Substantial effective sample sizes were required for external validation studies of predictive logistic regression models. *Journal of clinical epidemiology*. 2005;58(5):475-483.
- [141] Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. *Statistics in medicine*. 1997;16(9):965-980.
- [142] Moons KG, Altman DG, Reitsma JB, et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD): explanation and elaboration. *Annals of internal medicine*. 2015;162(1):W1-73.
- [143] Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. *Bmj*. 2009;338:b605.
- [144] Austin PC, Steyerberg EW. Graphical assessment of internal and external calibration of logistic regression models by using loess smoothers. *Statistics in medicine*. 2014;33(3):517-535.
- [145] Steyerberg EW, Vergouwe Y. Towards better clinical prediction models: seven steps for development and an ABCD for validation. *European heart journal*. 2014;35(29):1925-1931.
- [146] Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *Bmj*. 2003;326(7379):41-44.
- [147] Gold R, Reichman M, Greenberg E, et al. Developing a new reference standard: is validation necessary? *Academic radiology*. 2010;17(9):1079-1082.
- [148] Hawkins DM, Garrett JA, Stephenson B. Some issues in resolution of diagnostic tests using an imperfect gold standard. *Statistics in medicine*. 2001;20(13):1987-2001.
- [149] Ferrat E, Paillaud E, Caillet P, et al. Performance of Four Frailty Classifications in Older Patients With Cancer: Prospective Elderly Cancer Patients Cohort Study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2017;35(7):766-777.
- [150] Ferrat E, Audureau E, Paillaud E, et al. Four Distinct Health Profiles in Older Patients With Cancer: Latent Class Analysis of the Prospective ELCAPA Cohort. *J Gerontol A Biol Sci Med Sci*. 2016;71(12):1653-1660.
- [151] Raynaud-Simon A, Revel-Delhom C, Hebuterne X, French N, Health Program FHHA. Clinical practice guidelines from the French Health High Authority: nutritional support strategy in protein-energy malnutrition in the elderly. *Clin Nutr*. 2011;30(3):312-319.
- [152] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*. 1988;44(3):837-845.
- [153] Kenis C, Schuermans H, Van Cutsem E, et al. P8 Screening for a geriatric risk profile in older cancer patients: a comparative study of the predictive validity of three screening tools. *Critical Reviews in Oncology/Hematology*. 2009;72(1, Supplement):S22.
- [154] Luce S, De Breucker S, Van Gossum A, et al. How to identify older patients with cancer who should benefit from comprehensive geriatric assessment? *J Geriatr Oncol*. 2012;3(4):351-358.
- [155] Pialoux T, Goyard J, Lesourd B. Screening tools for frailty in primary health care: a systematic review. *Geriatrics & gerontology international*. 2012;12(2):189-197.
- [156] Nunes DP, Duarte YA, Santos JL, Lebrao ML. Screening for frailty in older adults using a self-reported instrument. *Revista de saude publica*. 2015;49:2.
- [157] Martinez-Tapia C, Paillaud E, Liuu E, et al. Prognostic value of the G8 and modified-G8 screening tools for multidimensional health problems in older patients with cancer. *European journal of cancer*. 2017;83:211-219.

- [158] Rockwood K, Mitnitski A. Frailty in relation to the accumulation of deficits. *J Gerontol A Biol Sci Med Sci*. 2007;62(7):722-727.
- [159] van Putten W. CART: Stata module to perform Classification and Regression Tree analysis. *Statistical Software Components, S456776* 2006; <https://ideas.repec.org/c/boc/bocode/s456776.html>. Accessed 08-17-2016.
- [160] Harrell FE, Jr., Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *Jama*. 1982;247(18):2543-2546.
- [161] Gönen M, Heller G. Concordance probability and discriminatory power in proportional hazards regression. *Biometrika*. 2005;92(4):965-970.
- [162] Benyamini Y, Blumstein T, Lusky A, Modan B. Gender differences in the self-rated health-mortality association: is it poor self-rated health that predicts mortality or excellent self-rated health that predicts survival? *Gerontologist*. 2003;43(3):396-405; discussion 372-395.
- [163] Cesari M, Onder G, Zamboni V, et al. Physical function and self-rated health status as predictors of mortality: results from longitudinal analysis in the iLSIRENTE study. *BMC geriatrics*. 2008;8:34.
- [164] Wang C, Satiriano WA. Self-rated current and future health independently predict subsequent mortality in an aging population. *J Gerontol A Biol Sci Med Sci*. 2007;62(12):1428-1434.
- [165] Koyi H, Hillerdal G, Andersson O, et al. Chemotherapy Treatment of Elderly Patients ( $\geq 70$  Years) with Non-Small Cell Lung Cancer: A Seven-Year Retrospective Study of Real-Life Clinical Practice at Karolinska University Hospital, Sweden. *Lung Cancer International*. 2015;2015:6.
- [166] Rades D, Conde-Moreno AJ, Segedin B, Veninga T, Cacicedo J, Schild SE. A Prognostic Instrument to Estimate the Survival of Elderly Patients Irradiated for Metastatic Epidural Spinal Cord Compression From Lung Cancer. *Clinical Lung Cancer*. 2015.
- [167] Verweij NM, Schiphorst AH, Pronk A, van den Bos F, Hamaker ME. Physical performance measures for predicting outcome in cancer patients: a systematic review. *Acta oncologica*. 2016;55(12):1386-1391.
- [168] Owusu C, Koroukian SM, Schluchter M, Bakaki P, Berger NA. Screening older cancer patients for a Comprehensive Geriatric Assessment: A comparison of three instruments. *J Geriatr Oncol*. 2011;2(2):121-129.
- [169] Misset JL, Bauer C. What is an "Elderly" oncologic patient? *Critical Reviews in Oncology / Hematology*. 2008;67(1):62-63.
- [170] Tucci A, Ferrari S, Bottelli C, Borlenghi E, Drera M, Rossi G. A comprehensive geriatric assessment is more effective than clinical judgment to identify elderly diffuse large cell lymphoma patients who benefit from aggressive therapy. *Cancer*. 2009;115(19):4547-4553.
- [171] Pottel L, Lycke M, Boterberg T, et al. G-8 indicates overall and quality-adjusted survival in older head and neck cancer patients treated with curative radiochemotherapy. *BMC cancer*. 2015;15:875.
- [172] Dubruille S, Libert Y, Roos M, et al. Identification of clinical parameters predictive of one-year survival using two geriatric tools in clinically fit older patients with hematological malignancies: Major impact of cognition. *J Geriatr Oncol*. 2015;6(5):362-369.
- [173] Gooiker GA, Dekker JW, Bastiaannet E, et al. Risk factors for excess mortality in the first year after curative surgery for colorectal cancer. *Annals of surgical oncology*. 2012;19(8):2428-2434.

- [174] Dekker JW, van den Broek CB, Bastiaannet E, van de Geest LG, Tollenaar RA, Liefers GJ. Importance of the first postoperative year in the prognosis of elderly colorectal cancer patients. *Annals of surgical oncology*. 2011;18(6):1533-1539.
- [175] Papis D, Lembo R, O'Reilly D, et al. The predictive ability of timed "Up & Go", VES-13 and GFI in hepato-pancreato-biliary onco-geriatric surgical patients. *J Geriatr Oncol*. 2012;3, Supplement 1:S100-S101.
- [176] Fried LP, Walston, J. Frailty and failure to thrive. In: Hazzard WR, Blass, J.P., Ettinger, W.H.Jr., Halter, J.B. and Ouslander, J., ed. *Principles of Geriatric Medicine and Gerontology, 4th Edition*. New York: McGraw Hill; 1998:1387-1402.
- [177] Mion LC, Palmer RM, Anetzberger GJ, Meldon SW. Establishing a case-finding and referral system for at-risk older individuals in the emergency department setting: the SIGNET model. *Journal of the American Geriatrics Society*. 2001;49(10):1379-1386.
- [178] Fargeas JB, Picat MA, Dumazeau L, et al. Validation of a geriatric screening tool for patients over 70 years old with malignant haemopathy undertaken in the haematologic network of Limousin. *Crit Rev Oncol Hemat*. 2009;72:S23-S24.
- [179] Soubeyran P, Rainfray M, Mathoulin-Pelissier S, Ford AB. Prediction of early death risk in the elderly with cancer: results of a prospective multicentric study of 364 patients under chemotherapy. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2007;ASCO. Annual Meeting Proceedings Part 1(25: 9040).
- [180] Denewet N, De Breucker S, Luce S, Kennes B, Higuët S, Pepersack T. Comprehensive geriatric assessment and comorbidities predict survival in geriatric oncology. *Acta Clin Belg*. 2016;71(4):206-213.
- [181] Aguayo GA, Donneau AF, Vaillant MT, et al. Agreement Between 35 Published Frailty Scores in the General Population. *American journal of epidemiology*. 2017;186(4):420-434.
- [182] Bellera CA, Artaud F, Rainfray M, Soubeyran PL, Mathoulin-Pelissier S. Modeling individual and relative accuracy of screening tools in geriatric oncology. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2017;28(5):1152-1157.
- [183] Soubeyran P, Terret C, Bellera C, et al. Role of geriatric intervention in the treatment of older patients with cancer: rationale and design of a phase III multicenter trial. *BMC cancer*. 2016;16(1):932.
- [184] Abellan Van Kan G, Sinclair A, Andrieu S, et al. The Geriatric Minimum Data Set for clinical trials (GMDS). *The journal of nutrition, health & aging*. 2008;12(3):197-200.
- [185] Etchepare F, Bellera C, Caillet P, et al. Mini data-set : les informations minimales pour l'évaluation gériatrique dans les essais cliniques conduits chez les patients âgés atteints de cancer. *Revue d'Épidémiologie et de Santé Publique*. 2017;65, Supplement 2:S60-S61.
- [186] Townsley CA, Selby R, Siu LL. Systematic review of barriers to the recruitment of older patients with cancer onto clinical trials. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2005;23(13):3112-3124.
- [187] Lewis JH, Kilgore ML, Goldman DP, et al. Participation of patients 65 years of age or older in cancer clinical trials. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2003;21(7):1383-1389.
- [188] Scher KS, Hurria A. Under-representation of older adults in cancer registration trials: known problem, little progress. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2012;30(17):2036-2038.

- [189] Brunello A, Fontana A, Zafferri V, et al. Development of an oncological-multidimensional prognostic index (Onco-MPI) for mortality prediction in older cancer patients. *Journal of cancer research and clinical oncology*. 2016;142(5):1069-1077.
- [190] Luciani A, Ascione G, Bertuzzi C, et al. Detecting disabilities in older patients with cancer: comparison between comprehensive geriatric assessment and vulnerable elders survey-13. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2010;28(12):2046-2050.
- [191] Biganzoli L, Boni L, Becheri D, et al. Evaluation of the cardiovascular health study (CHS) instrument and the Vulnerable Elders Survey-13 (VES-13) in elderly cancer patients. Are we still missing the right screening tool? *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2013;24(2):494-500.
- [192] Monfardini S, Basso U, Fiduccia P, et al. Can the short screening test Vulnerable Elders Survey 13 (VES-13) substitute for the time-consuming comprehensive geriatric assessment (CGA) to identify vulnerable/frail elderly breast cancer patients? *Journal of Clinical Oncology*. 2010;28(15\_suppl):9114-9114.
- [193] Falci C, Basso U, Fiduccia P, et al. Is vulnerable elders survey 13 (VES-13) a sensitive and specific screening tool for identifying vulnerable/frail elderly cancer patients compared to full comprehensive geriatric assessment (CGA)? *Crit Rev Oncol Hematol*. 2009;72:S19-S20.
- [194] Kellen E, Bulens P, Deckx L, et al. Identifying an accurate pre-screening tool in geriatric oncology. *Crit Rev Oncol Hematol*. 2010;75(3):243-248.
- [195] Molina-Garrido MJ G-PC, Munoz Sanchez M, Garcia-Cano J. A comparison of the fried criteria and the VES-13 questionnaire as screening instruments for frail elderly patients diagnosed with cancer. *European Geriatric Medicine*. 2011;2(suppl 1):34 (abstr PC-030).
- [196] Mohile SG, Bylow K, Dale W, et al. A pilot study of the vulnerable elders survey-13 compared with the comprehensive geriatric assessment for identifying disability in older patients with prostate cancer who receive androgen ablation. *Cancer*. 2007;109(4):802-810.
- [197] Molina-Garrido MJ, Guillen-Ponce C. Comparison of two frailty screening tools in older women with early breast cancer. *Crit Rev Oncol Hematol*. 2011;79(1):51-64.
- [198] Luciani A, Dottorini L, Battisti N, et al. Screening elderly cancer patients for disabilities: evaluation of study of osteoporotic fractures (SOF) index and comprehensive geriatric assessment (CGA). *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO*. 2013;24(2):469-474.
- [199] Kristjansson SR, Jordhoy MS, Nesbakken A, Wyller TB. A comparison of two methods to measuring frailty in elderly patients with colorectal cancer. *Critical Reviews in Oncology / Hematology*.68:S30.
- [200] Molina-Garrido MJ, Guillen-Ponce C. Validation of the Barber Test as a screening tool of frailty in elderly cancer patients. Preliminary results. *J Geriatr Oncol*.3:S76-S77.
- [201] Fargeas JB, Picat MA, Duzmazeau L, et al. Validation of a geriatric screening tool for patients over 70 years old with malignant haemopathy undertaken in the hematologic

**Annex 1. Studies comparing screening tools with Geriatric Assessment  
in older patients with cancer**

**Table A1.1.** Diagnostic performances of screening tools against geriatric assessment in older patients with cancer.

Screening tool	No. of patients	Type of cancer	No. of GA items	≥1 impairment		≥2 impairments	
				Se (%)	Sp (%)	Se (%)	Sp (%)
<b>G8</b>							
Bellera [61]	364	Various	7	85	65		
Luce [154]	211	Various	10	65	3		
Velghe [71]	50	Hematological	6	89	100		
Soubeyran [85]	1435	Various	7	77	64	87	55
Liuu [121]	518	Various	7	87	60	93	53
Pamoukdjian [75]	269	Various	6	90	35	93	29
Kenis [153]	140	Various	7			80	40
Baitar [120]	170	Various	8			92	52
Kenis [88]	937	Various	7			87	59
Pottel [86]	51	Head and neck	7			86	75
Smets [68]	108	Various	5			87	64
Hamaker [67]	108	Hematological	8			69	79
Holmes [70]	50	Hematological	8			70	54
Kenig [73]	135	Abdominal tumors	8			97	44
Hentschel [74]	84	Various	6			38	62
<b>VES-13</b>							
Luciani [190]	419	Various	8	87	62		
Biganzoli [191]	259	Various	5	62	81		
Monfardini [192]	150	Breast	4 <sup>a</sup>	68	71		
Falci [193]	93	Various	4 <sup>a</sup>	60	70		
Soubeyran [85]	1435	Various	7	69	74	79	64
Owusu [168]	117	Various	10			88	69
Kellen [194]	113	Various	5			61	78
Molina-Garrido [195]	58	Various	7			39	100
Pottel [86]	51	Head and neck	7			57	100
Mohile [196]	50	Prostate	7			73	86
Molina-Garrido [197]	41	Breast	7			55	100
Hentschel [74]	84	Various	6			57	79
Holmes [70]	50	Haematological	8			15	100
Smets [68]	108	Various	5			67	70
Augschoell [72]	76	Various	10			56	91
Kenig [69]	64	Various	8			60	78
Kenig [73]	135	Abdominal tumors	8			69	81
<b>fTRST</b>							
Kenis [88] C/O 1	937	Various	7			91	42
Kenis [88] C/O 2	937	Various	7			67	80
Kenis [153] C/O 1	140	Various	7			92	50
Kenis [153] C/O 2	140	Various	7			64	100
Kenig [73]	135	Abdominal tumors	8			59	86

<b>GFI</b>						
Baitar [120]	170	Various	8		66	87
Kenis [153]	140	Various	7		57	87
Kellen [194]	113	Various	5		39	86
Smets [68]	108	Various	5		79	71
Kenig [73]	135	Abdominal tumors	8		64	86
<b>SOF</b>						
Luciani [198]	400	Various	6		89	81
<b>KPS</b>						
Luce [154]	211	Various	10	29	44	
Owusu [168]	117	Various	10			78 91
<b>Fried criteria</b>						
Biganzoli [191] C/O 1	259	Various	5	87	49	
Biganzoli [191]	259	Various	5	31 <sup>b</sup>	98 <sup>b</sup>	
Kristjansson [199]	74	Colorectal	6	25	96	
Molina-Garrido [195]	58	Various	7			37 86
Kenig [73]	135	Abdominal tumors	8			52 92
<b>Barber</b>						
Molina-Garrido [200]	173	Various	7		74	39
Molina-Garrido [197]	41	Breast	7		59	79
<b>ISAR</b>						
Luce [154]	211	Various	10	70	10	
<b>OGS</b>						
Valéro [64]	126	Various	5	88	44	
<b>ECOG-PS</b>						
Owusu [168]	117	Various	10			94 55
Kenig [73]	135	Abdominal tumors	8			54 81
<b>aCGA</b>						
Kellen [194]	113	Various	5		51	97
Smets [68]	108	Various	5		79	59
Kenig [73]	135	Abdominal tumors	8		84	86
<b>Gerhematolim</b>						
Fargeas [201]	104	Hematological	8		95	87
<b>SAOP2</b>						
Extermann [83]	31	Various	7	100	40	

Abbreviations: GA, geriatric assessment; Se, sensitivity; Sp, specificity; C/O, Cut-off

<sup>a</sup> Balducci classification (Fit vs. Non-fit). Domains: age, activities of daily living, comorbidities and geriatric syndromes.

<sup>b</sup> Approximated values based on the area under the ROC curve



## **Annex 2. Geriatric Assessment Tools**

## Geriatric Assessment

<b>IADL – Instrumental Activities of Daily Living</b>
<p><b><u>Ability to use telephone</u></b></p> <p>1 Operates telephone on own initiative; looks up and dials numbers            1 Dials a few well-known numbers            1 Answers telephone, but does not dial            0 Does not use telephone at all</p>
<p><b><u>Shopping</u></b></p> <p>1 Takes care of all shopping needs independently            0 Shops independently for small purchases            0 Needs to be accompanied on any shopping trip            0 Completely unable to shop</p>
<p><b><u>Food preparation</u></b></p> <p>1 Plans, prepares, and serves adequate meals independently            0 Prepares adequate meals if supplied with ingredients            0 Heats and serves prepared meals or prepares meals but does not maintain adequate diet            0 Needs to have meals prepared and served</p>
<p><b><u>Housekeeping</u></b></p> <p>1 Maintains house alone with occasion assistance (heavy work)            1 Performs light daily tasks such as dishwashing, bed making            1 Performs light daily tasks, but cannot maintain acceptable level of cleanliness            1 Needs help with all home maintenance tasks            0 Does not participate in any housekeeping tasks</p>
<p><b><u>Laundry</u></b></p> <p>1 Does personal laundry completely            1 Launders small items, rinses socks, stockings, etc.            0 All laundry must be done by others</p>
<p><b><u>Mode of transportation</u></b></p> <p>1 Travels independently on public transportation or drives own car            1 Arranges own travel via taxi, but does not otherwise use public transportation            1 Travels on public transportation when assisted or accompanied by another            0 Travel limited to taxi or automobile with assistance of another            0 Does not travel at all</p>
<p><b><u>Responsibility for own medications</u></b></p> <p>1 Is responsible for taking medication in correct dosages at correct time            0 Takes responsibility if medication is prepared in advance in separate dosages            0 Is not capable of dispensing own medication</p>
<p><b><u>Ability to handle finances</u></b></p> <p>1 Manages financial matters independently (budgets, writes checks, pays rent and bills, goes to bank); collects and keeps track of income            1 Manages day-to-day purchases, but needs help with banking, major purchases, etc.            0 Incapable of handling money</p>
<b>Total Score IADL __/8</b>

<b>ADL – Activities of Daily Living</b>	
<b>Bathing</b>	
1 Receives no assistance	
½ Receives assistance in bathing only one part of the body ( such as back or a leg)	
0 Receives assistance in bathing more than one part of the body (or not bathed)	
<b>Dressing</b>	
1 Receives no assistance	
½ Receives assistance in bathing only one part of the body ( such as back or a leg)	
0 Receives assistance in bathing more than one part of the body (or not bathed)	
<b>Toileting</b>	
1 Receives no assistance	
½ Receives assistance in bathing only one part of the body ( such as back or a leg)	
0 Receives assistance in bathing more than one part of the body (or not bathed)	
<b>Transfer</b>	
1 Receives no assistance	
½ Receives assistance in bathing only one part of the body ( such as back or a leg)	
0 Receives assistance in bathing more than one part of the body (or not bathed)	
<b>Continenence</b>	
1 Receives no assistance	
½ Receives assistance in bathing only one part of the body ( such as back or a leg)	
0 Receives assistance in bathing more than one part of the body (or not bathed)	
<b>Feeding</b>	
1 Receives no assistance	
½ Receives assistance in bathing only one part of the body ( such as back or a leg)	
0 Receives assistance in bathing more than one part of the body (or not bathed)	
<b>Total score ADL ___/6</b>	

<b>TUG – Timed Get Up and Go</b>	
<input type="checkbox"/> Not feasible (The patient did the test = 0 ; The patient did not do the test = 1)	
On the command “go”, the patient: <ul style="list-style-type: none"> <li>• Stand up from the chair _____</li> <li>• Walk to the line on the floor at your normal pace (3m) _____</li> <li>• Turn around and walk back to the chair (3m) _____</li> <li>• Sit down _____</li> </ul> Score ____/ 4	Time: ____seconds OR <input type="checkbox"/> ≤ 20 sec. <input type="checkbox"/> > 20sec
<b>Risk of falls</b> (if score ≥ 3 and/or time > 20 sec.) <input type="checkbox"/> No <input type="checkbox"/> YES	

<b>Mini-GDS – Geriatric Depression Scale 4 items (Used in ELCAPA – Development cohort)</b>		
1. Do you often feel downhearted and blue?	<b>Yes = 1</b>	No = 0
2. Do you feel that your life is empty?	<b>Yes = 1</b>	No = 0
3. Do you feel happy most of the time?	Yes = 0	<b>No = 1</b>
4. Do you feel that your situation is hopeless?	<b>Yes = 1</b>	No = 0
<b>Score</b>		____ / 4

<b>GDS-15 – Geriatric Depression Scale 15 items (Used in ONCODAGE – Validation cohort)</b>		
1. Are you basically satisfied with your life?	<b>Yes = 1</b>	No = 0
2. Have you dropped many of your activities and interests?	<b>Yes = 1</b>	No = 0
3. Do you feel that your life is empty?	<b>Yes = 1</b>	No = 0
4. Do you often get bored?	<b>Yes = 1</b>	No = 0
5. Are you in good spirits most of the time?	Yes = 0	<b>No = 1</b>
6. Are you afraid that something bad is going to happen to you?	<b>Yes = 1</b>	No = 0
7. Do you feel happy most of the time?	Yes = 0	<b>No = 1</b>
8. Do you often feel helpless?	<b>Yes = 1</b>	No = 0
9. Do you prefer to stay at home, rather than going out and doing new things?	<b>Yes = 1</b>	No = 0
10. Do you feel you have more problems with memory than most?	<b>Yes = 1</b>	No = 0
11. Do you think it is wonderful to be alive now?	Yes = 0	<b>No = 1</b>
12. Do you feel pretty worthless the way you are now?	<b>Yes = 1</b>	No = 0
13. Do you feel full of energy?	Yes = 0	<b>No = 1</b>
14. Do you feel that your situation is hopeless?	<b>Yes = 1</b>	No = 0
15. Do you think that most people are better off than you are?	<b>Yes = 1</b>	No = 0
<b>Score</b>	<b>___ / 15</b>	

<b>Score CIRS-G – Cumulative Illness Rating Scale – Geriatric</b>			
<b>The scoring system is as follows:</b>			
0. No problem affecting that system			
1. Current mild problem or past significant problem			
2. Moderate disability or morbidity/ requires "first line" therapy			
3. Severe/constant significant disability/ "uncontrollable" chronic problems			
4. Extremely Severe/immediate treatment required/end organ failure/severe impairment in function			
✎ If there are several diseases in the same system, score the most severe.			
<b>Heart</b> Heart only		<b>Liver</b> Liver only	
<b>Vascular / Hypertension</b>		<b>Renal</b> Kidneys only	
<b>Hematopoietic</b> Blood, blood vessels and cells, marrow, spleen, lymphatic		<b>Genitourinary</b> Ureters, bladder, urethra, prostate, genitals, uterus, ovaries	
<b>Respiratory</b> Lungs, bronchi, trachea below the larynx		<b>Musculoskeletal and teguments</b> Muscles, bone and skin	
<b>Eyes, ears, nose and throat and larynx</b>		<b>Neurological</b> Brain, spinal cord and nerves	
<b>Upper gastrointestinal tract</b> esophagus, stomach, duodenum, biliary trees, pancreas		<b>Endocrine-metabolic</b> Diabetes, diffuse infections and poisonings, nutrition	
<b>Lower gastrointestinal tract</b> Intestines, hernias		<b>Psychiatric illness</b> Dementia, agitation, depression, anxiety, psychoses	
<b>(A) Total score</b>			
<b>(B) Total number of categories endorsed</b>			
Severity index = A/B			
Number of <b>categories at level 3</b>			
Number of <b>categories at level 4</b>			

## MNA – Mini Nutritional Assessment

Screening and G8	Score
<b>Has food intake declined over the past 3 months due to loss of appetite, digestive problems, chewing or swallowing difficulties?</b> Anorexia <b>0</b> – severe <b>1</b> – moderate <b>2</b> – no anorexia	
<b>Weight loss during the last 3 months?</b> <b>0</b> – > 3 kg <b>1</b> – does not know <b>2</b> – between 1 and 3 kg <b>3</b> – no weight loss	
<b>Mobility</b> <b>0</b> – bed or chair bound <b>1</b> – able to get out of bed / chair but does not go out <b>2</b> – goes out	
<b>Has suffered psychological stress or acute disease in the past 3 months?</b> <b>0</b> – yes <b>2</b> – no	
<b>Neuropsychological problems? Dementia or depression</b> <b>0</b> – severe <b>1</b> – moderate <b>2</b> – no psychological problems	
<b>Body Mass Index (BMI) = weight in kg / (height in m)<sup>2</sup></b> <b>0</b> – IMC < 19 <b>1</b> – 19 ≤ IMC < 21 <b>2</b> – 21 ≤ IMC < 23 <b>3</b> – IMC ≥ 23	
<b>Screening score</b>	<b>/ 14</b>
Assessment	Score
<b>Lives independently (not in nursing home or hospital)?</b> <b>1</b> – Yes <b>0</b> – No	
<b>Takes more than 3 prescription drugs per day?</b> <b>0</b> – Yes <b>1</b> – No	
<b>Pressure sores or skin ulcers?</b> <b>0</b> – Yes <b>1</b> – No	
<b>How many full meals does the patient eat daily?</b> <b>0</b> – 1 meal <b>1</b> – 2 meals <b>2</b> – 3 meals	
<b>Selected consumption markers for protein intake</b> At least one serving of dairy products (milk, cheese, yoghurt) per day      Yes – No Two or more servings of legumes or eggs per week      Yes – No Meat, fish or poultry every day      Yes – No <b>0</b> – if 0 or 1 yes <b>0.5</b> – if 2 yes <b>1</b> – if 3 yes	
<b>Consumes two or more servings of fruit or vegetables per day?</b> <b>1</b> – Yes <b>0</b> – No	
<b>How much fluid (water, juice, coffee, tea, milk...) is consumed per day?</b> <b>0</b> – less than 3 cups <b>0.5</b> – 3 to 5 cups <b>1</b> – more than 5 cups	
<b>Mode of feeding</b> <b>0</b> – unable to eat without assistance <b>1</b> – self-fed with some difficulty <b>2</b> – self-fed without any problem	
<b>Self-view of nutritional status (Malnutrition)</b> <b>0</b> – severe malnutrition <b>1</b> – is uncertain or moderate malnutrition <b>2</b> – no nutritional problem	
<b>In comparison with other people of the same age, how does the patient consider his / her health status?</b> <b>0</b> – not as good <b>0.5</b> – does not know <b>1</b> – as good <b>2</b> – better	
<b>Mid-arm circumference (MAC) in cm</b> <b>0</b> – MAC < 21 cm <b>0.5</b> – 21 cm ≤ MAC ≤ 22 cm <b>1</b> – MAC > 22 cm	
<b>Calf circumference (CC) in cm</b> <b>0</b> – CC < 31 cm <b>1</b> – CC ≥ 31 cm	
<b>Assessment score</b>	<b>/ 16</b>
<b>* SCORE MNA TOTAL</b>	
<b>/ 30</b>	
<b>Interpretation</b> <input type="checkbox"/> Normal nutritional status      > 23.5 / 30 <input type="checkbox"/> At risk of malnutrition      17 to 23.5 / 30 <input type="checkbox"/> Malnourished      < 17 / 30	
<b>Age</b> <b>0</b> – > 85 years <b>1</b> – Between 80-85 years <b>2</b> – < 80 years	

## MMSE – Mini Mental State Examination

### Orientation \_\_\_\_\_ / 10

“What is today’s full date?”

- |    |  |   |
|----|--|---|
| 1. | Year .....   | 1 |
| 2. | Season (1 week in advance for the next season, 2 weeks late for the past season) ..... | 1 |
| 3. | Month ( $\pm$ 1 day) .....   | 1 |
| 4. | Day of the month ( $\pm$ 1 day) .....  | 1 |
| 5. | Day of the week .....  | 1 |

“Where are we now?”

- |     |                  |   |
|-----|------------------|---|
| 6.  | Country .....    | 1 |
| 7.  | Town .....       | 1 |
| 8.  | District .....   | 1 |
| 9.  | Hospital .....   | 1 |
| 10. | Ward/Floor ..... | 1 |

### Registration \_\_\_\_\_ / 3

The examiner names three objects and asks the patient to repeat all three of them; the patient learns the 3 names repeating until correct for a later question.

- |     |       |    |        |    |       |   |
|-----|-------|----|--------|----|-------|---|
| 11. | Lemon | or | Cigar  | or | Chair | 1 |
| 12. | Key   | or | Flower | or | Tulip | 1 |
| 13. | Ball  | or | Door   | or | Duck  | 1 |

Repeat the 3 names

(1 point for each correct the first time. If not, the examiner repeats them until patient learns all of them.)

### Attention and Calculation \_\_\_\_\_ / 5

“I would like you to subtract 7 from 100. Continue five times”

(In case of mistake, say: « Are you sure?”

- |     |          |   |
|-----|----------|---|
| 14. | 93 ..... | 1 |
| 15. | 86 ..... | 1 |
| 16. | 79 ..... | 1 |
| 17. | 72 ..... | 1 |
| 18. | 65 ..... | 1 |

**For all patients ask:**

Spell **WORLD** backwards  $\Rightarrow$  D-L-R-O-W

### Recall \_\_\_\_\_ / 3

“Earlier I told you the names of three things. Can you tell me what those were?”

(Score 1 point for each word correctly repeated at first trial.)

- |     |       |    |        |    |       |   |
|-----|-------|----|--------|----|-------|---|
| 19. | Lemon | or | Cigar  | or | Chair | 1 |
| 20. | Key   | or | Flower | or | Tulip | 1 |
| 21. | Ball  | or | Door   | or | Duck  | 1 |

### Language \_\_\_\_\_ / 8

- |     |   |   |
|-----|---|---|
| 22. | “Show the patient a pencil. « What is the name of this object?” ..... | 1 |
| 23. | “Show the patient a watch. « What is the name of this object?” .....  | 1 |
| 24. | “Repeat the phrase: No ifs, ands, or buts” .....                      | 1 |

**The examiner gives the patient a piece of blank paper and asks:**

- |     |   |   |
|-----|---|---|
| 25. | “Take the paper in your right hand, ..... | 1 |
| 26. | fold it in half .....                     | 1 |
| 27. | and put it on the floor” .....            | 1 |

(Score 1 point for each stage correctly done)

**Ask the patient to read and obey a written command on a piece of paper. The written instruction is “Close your eyes”:**

- |     |  |   |
|-----|--|---|
| 28. | “Please read this and do what it says” ..... | 1 |
|-----|--|---|

**Ask the patient to write a complete sentence:**

- |     |   |   |
|-----|---|---|
| 29. | “Make up and write a complete sentence about anything you like” ..... | 1 |
|-----|---|---|

(Score 1 if it is sensible and has a subject and a verb. Allow 30 seconds).

### Copying \_\_\_\_\_ / 1

**The examiner gives the patient a blank piece of paper and asks:**

- |     |                                  |   |
|-----|----------------------------------|---|
| 30. | «Please copy this picture” ..... | 1 |
|-----|----------------------------------|---|

(Score 1 point if every angle is present and if the figure is in both sides of the paper. We can allow several trials and allow 1 minute).

MMSE – Score	____ / 30
--------------	-----------

**Annex 3. Comparisons of baseline characteristics of included and non-included patients in each analysis**

**Table A3.1.** Patient characteristics comparisons between included and non-included patients in the screening tool development analysis.

Characteristics	Included population N=729	Not-included population N=327	p-value <sup>a</sup>
	N (%)	N (%)	
Age in years, median (IQR)	80 (76 ;84)	80 (76 ;85)	0.832
Male gender	387 (53.1)	155 (47.4)	0.087
Living alone at home	274 (37.6)	121 (37.6)	0.985
Cancer site			
Breast	137 (18.8)	41 (12.5)	<b>&lt;0.0001</b>
Colorectal	131 (18.0)	95 (29.1)	
Upper digestive tract and liver	118 (16.2)	69 (21.1)	
Urinary tract	117 (16.0)	29 (8.9)	
Prostate	99 (13.6)	20 (6.1)	
Hematological	49 (6.7)	36 (11.0)	
Other <sup>b</sup>	78 (10.7)	37 (11.3)	
Metastasis	299 (41.0)	140 (42.8)	0.292
Number of medications per day, median (IQR)	6 (4 ;8)	6 (4 ;8)	0.060
ECOG-PS			<b>&lt;0.0001</b>
0 - Fully active	177 (24.3)	43 (13.2)	
1 - Restricted activity but ambulatory	213 (29.3)	74 (22.8)	
2 - Up >50% of waking hours	129 (17.7)	54 (16.6)	
3 - Confined to bed >50% of the day	148 (20.3)	95 (29.2)	
4 - Completely disabled	61 (8.4)	59 (18.2)	
Geriatric assessment parameters			
ADL≤5	218 (29.9)	32/84 (38.1)	0.123
IADL≤7	457 (62.7)	53/80 (66.3)	0.531
MMSE≤23	193 (26.5)	10/36 (27.8)	0.867
Mini-GDS≥1	250 (34.3)	22/83 (26.5)	0.154
MNA≤23.5	426 (58.4)	45/79 (57.0)	0.801
CIRS-G≥1 comorbidity grade 3/4	414 (56.8)	39/60 (65.0)	0.216
TUG≥20 s	304 (41.7)	42/81 (51.9)	0.080

Abbreviations: IQR, interquartile range; ECOG-PS, Eastern Cooperative Oncology Group-Performance Status; ADL, Activities of Daily Living; IADL, Instrumental ADL; MMSE, Mini Mental State Examination; Mini-GDS, Mini Geriatric Depression Scale; MNA, Mini Nutritional Assessment; CIRS-G, Cumulative Illness Rating Scale for Geriatrics; TUG, Timed Up-and-Go test

<sup>a</sup> p-value: t-test or Wilcoxon Mann-Whitney test for quantitative variables and chi-square or Fischer's exact test for qualitative variables, as appropriate.

<sup>b</sup> Included/Non-included population: unknown primary origin (n=21/11), lung (n=17/4), skin (n=14/5), sarcoma (n=9/6), brain (n=5/1), gynecologic (n=4/4), others (n=8/6)

**Table A3.2.** Patient characteristics comparisons between included and non-included patients in the external validation analysis

Characteristics	Included population	Not-included population	p-value <sup>a</sup>
	N=1304	N=131	
	N (%)	N (%)	
Age in years, median (IQR)	78 [74-82]	77 [74-81]	0.332
Male gender	406 (31.1)	28 (21.4)	<b>0.020</b>
Living alone at home	560 (42.9)	61 (46.6)	0.330
Cancer site			0.118
Breast	688 (52.8)	86 (65.6)	
Colorectal	191 (14.6)	13 (9.9)	
Prostate	112 (8.6)	10 (7.6)	
Hematological	103 (7.9)	9 (6.9)	
Lung	140 (10.7)	9 (6.9)	
Upper aerodigestive tract	70 (5.4)	4 (3.1)	
Metastasis (n=1202)	183 (16.8)	14 (12.7)	0.208
Number of medications per day, median (IQR)	5 [3-7]	5 [3-7]	0.074
ECOG-PS (n=1352)			<b>0.028</b>
0 - Fully active	523 (40.1)	26 (54.2)	
1 - Restricted activity but ambulatory	457 (35.0)	19 (39.6)	
2 - Up >50% of waking hours	198 (15.2)	1 (2.1)	
3 - Confined to bed >50% of the day	81 (6.2)	1 (2.1)	
4 - Completely disabled	45 (3.5)	1 (2.1)	
Abnormal Geriatric Assessment (n=1397)	1035 (79.4)	74 (79.6)	0.963
ADL≤5	204 (15.6)	12 (9.4)	
IADL≤7	600 (46.8)	52 (43.7)	
MMSE≤23	253 (19.5)	28 (21.9)	
GDS-15≥6	390 (30.7)	28 (23.1)	
MNA≤23.5	595 (45.8)	36 (28.6)	
CIRS-G≥1 comorbidity grade 3/4	546 (42.2)	29 (25.7)	
TUG≥20 s	162 (13.8)	6 (5.8)	

Abbreviations: IQR, interquartile range; ECOG-PS, Eastern Cooperative Oncology Group-Performance Status; ADL, activities of daily living; IADL, Instrumental ADL; MMSE, Mini Mental State Examination; Mini-GDS, Mini Geriatric Depression Scale; MNA, Mini Nutritional Assessment; CIRS-G, Cumulative Illness Rating Scale for Geriatrics; TUG, timed up-and-go test.

<sup>a</sup> p-value: t-test or Wilcoxon Mann-Whitney test for quantitative variables and chi-square or Fischer's exact test for qualitative variables, as appropriate.

**Table A3.3.** Patient characteristics comparisons between included and non-included patients in the Gold standard definitions analysis.

Characteristics	Included population N=1136	Not-included population N=807	p-value <sup>a</sup>
	N (%)	N (%)	
Age in years, median (IQR)	80 (76;85)	80 (77;84)	0.472
Male gender	587 (51.7)	404 (50.2)	0.537
Cancer site			0.790
Breast	182 (16.0)	136 (16.9)	
Colorectal	201 (17.7)	154 (19.2)	
Upper digestive tract and liver	212 (18.7)	142 (17.7)	
Urinary tract	167 (14.7)	119 (14.8)	
Prostate	127 (11.2)	82 (10.2)	
Hematological	84 (7.4)	48 (6.0)	
Other <sup>b</sup>	163 (14.3)	123 (15.3)	
Metastasis (n=1565)	478/922 (51.8)	313/643 (48.7)	0.223
No. of drugs, median (IQR)	6 (4;8)	6 (4;8)	0.100
ECOG Performance Status			0.140
0 - Fully active	205 (18.0)	138 (17.6)	
1 - Restricted activity but ambulatory	342 (30.1)	232 (29.6)	
2 - Up >50% of waking hours	220 (19.4)	149 (19.0)	
3 - Confined to bed >50% of the day	260 (22.9)	161 (20.5)	
4 - Completely disabled	109 (9.6)	104 (13.3)	
ADL≤5	359 (31.6)	290/790 (36.7)	<b>0.020</b>
IADL≤7	665 (63.5)	474/664 (71.4)	<b>0.0008</b>
MMSE≤23	285 (28.3)	154/546 (28.2)	0.977
Mini-GDS≥1	379 (35.7)	211/661 (31.9)	0.112
MNA≤23.5	730 (65.9)	342/591 (57.9)	<b>0.001</b>
CIRS-G≥1 comorbidity grade 3/4	718 (63.2)	433/653 (66.3)	0.187
TUG≥20 s	434 (41.8)	262/660 (39.7)	0.397
Inappropriate social environment <sup>c</sup>	177 (15.6)	122/787 (84.5)	0.962
G8			0.129
Normal >14	147 (12.9)	115 (15.4)	
Abnormal ≤14	989 (87.1)	631 (84.6)	
Median (IQR)	10.5 (8;13.5)	11.5 (8.5;14)	<b>0.0003</b>
Modified G8			0.639
Normal <6	167 (14.7)	106 (13.9)	
Abnormal ≥6	969 (85.3)	655 (86.1)	
Median (IQR)	19 (10;25)	16 (8;25)	<b>0.012</b>

Abbreviations: IQR, interquartile range; ECOG, Eastern Cooperative Oncology Group; ADL, activities of daily living; IADL, Instrumental ADL; MMSE, Mini Mental State Examination; Mini-GDS, Mini Geriatric Depression Scale; MNA, Mini Nutritional Assessment; CIRS-G, Cumulative Illness Rating Scale for Geriatrics; TUG, timed up-and-go test.

<sup>a</sup> p-value: t-test or Wilcoxon Mann-Whitney test for quantitative variables and chi-square or Fischer's exact test for qualitative variables, as appropriate.

<sup>b</sup> Unknown primary origin (n=41), lung (n=47), skin (n=52), sarcoma (n=19), gynecologic (n=21), brain (n=11), head and neck (n=10), thyroid (n=5), others (n=14).

<sup>c</sup> defined as absence of a primary caregiver or of adequate support at home or of a strong circle of family and friends able to meet the needs of the patient at the time of the evaluation.

**Table A3.4.** Patient characteristics comparisons between included and non-included patients in the survival analysis.

Characteristics	Included population N=1333	Not-included population N=280	p-value <sup>a</sup>
	N (%)	N (%)	
Age in years, median (IQR)	80 (76;84)	81 (77;86)	<b>0.0002</b>
Male gender	690 (51.8)	128 (45.7)	0.066
Cancer site			<b>0.003</b>
Breast	218 (16.4)	45 (16.1)	
Colorectal	261 (19.6)	45 (16.1)	
Upper digestive tract and liver	228 (17.1)	67 (23.9)	
Urinary tract	197 (14.8)	34 (12.1)	
Prostate	155 (11.6)	18 (6.4)	
Hematological	105 (7.9)	20 (7.1)	
Other <sup>b</sup>	169 (12.7)	51 (18.2)	
Metastasis (n=1280)	524/1063 (49.3)	116/217 (53.5)	0.264
No. of drugs, median (IQR)	6 (4;8)	6 (4;9)	0.840
No. of comorbidities, median (IQR)	2 (1;3)	2 (1;3)	0.071
ECOG Performance Status		(n=267)	<b>0.002</b>
0 - Fully active	276 (20.7)	34 (12.7)	
1 - Restricted activity but ambulatory	383 (28.7)	79 (29.6)	
2 - Up >50% of waking hours	240 (18.0)	61 (22.9)	
3 - Confined to bed >50% of the day	295 (22.1)	51 (19.1)	
4 - Completely disabled	139 (10.4)	42 (15.7)	
Abnormal Geriatric Assessment (n=1519)	1170/1271 (92.1)	246/248 (99.2)	
ADL≤5	440 (33.1)	113 (41.7)	<b>0.007</b>
IADL≤7	763 (63.6)	148 (78.3)	<b>&lt;0.0001</b>
MMSE≤23	246 (25.1)	58 (34.3)	<b>0.012</b>
Mini-GDS≥1	404 (33.6)	76 (38.8)	0.158
MNA≤23.5	844 (63.8)	67 (56.3)	0.102
CIRS-G≥1 comorbidity grade 3/4	767 (62.5)	139 (64.1)	0.609
TUG≥20 s	519 (42.3)	79 (38.7)	0.350
Screening scores			
G8			
Normal >14	218 (16.4)	13 (7.0)	<b>0.0008</b>
Abnormal ≤14	1115 (83.6)	174 (93.0)	
Median (IQR)	11 (8;13.5)	10 (7;13)	<b>0.003</b>
Modified G8			
Normal <6	225 (16.9)	7 (8.8)	0.057
Abnormal ≥6	1108 (83.1)	73 (91.3)	
Median (IQR)	17 (9;25)	17 (9;25)	0.819

Abbreviations: IQR, interquartile range; ECOG, Eastern Cooperative Oncology Group; ADL, activities of daily living; IADL, Instrumental ADL; MMSE, Mini Mental State Examination; Mini-GDS, Mini Geriatric Depression Scale; MNA, Mini Nutritional Assessment; CIRS-G, Cumulative Illness Rating Scale for Geriatrics; TUG, timed up-and-go test.

<sup>a</sup> p-value: t-test or Wilcoxon Mann-Whitney test for quantitative variables and chi-square or Fischer's exact test for qualitative variables, as appropriate.

<sup>b</sup> Unknown primary origin (n=41), lung (n=47), skin (n=52), sarcoma (n=19), gynecologic (n=21), brain (n=11), head and neck (n=10), thyroid (n=5), others (n=14).



## **Annex 4. Univariate analysis according to cancer site**

**Table A4.1.** Univariate analysis of the association between G8 items and an abnormal Geriatric Assessment, by cancer site.

G8 items	Digestive (n=249)			Breast (n=137)			Urinary tract (n=118)			Prostate (n=99)			Hematological (n=49)		
	Geriatric Assessment			Geriatric Assessment			Geriatric Assessment			Geriatric Assessment			Geriatric Assessment		
	Normal	Abnormal <sup>a</sup>	P-value <sup>b</sup>	Normal	Abnormal <sup>a</sup>	P-value <sup>b</sup>	Normal	Abnormal <sup>a</sup>	P-value <sup>b</sup>	Normal	Abnormal <sup>a</sup>	P-value <sup>b</sup>	Normal	Abnormal <sup>a</sup>	P-value <sup>b</sup>
<b>Anorexia</b>															
Absent	15(83.3%)	81(35.1%)	<b>&lt;0.001</b>	28(90.3%)	65(61.3%)	<b>0.008</b>	12(92.3%)	57(54.8%)	<b>0.040</b>	27(100%)	41(56.9%)	<b>&lt;0.0001</b>	3(75.0%)	23(51.1%)	1.000
Moderate	1(5.6%)	57(24.7%)		3(9.7%)	36(34.0%)		1(7.7%)	34(32.7%)		0(0.0%)	21(29.2%)		0(0.0%)	6(13.3%)	
Severe	2(11.1%)	93(40.3%)		0(0.0%)	5(4.7%)		0(0.0%)	13(12.5%)		0(0.0%)	10(13.9%)		1(25.0%)	16(35.6%)	
<b>Weight loss</b>															
Absent	8(44.4%)	52(22.5%)	<b>0.005</b>	28(90.3%)	60(56.6%)	<b>0.002</b>	12(92.3%)	32(30.8%)	<b>&lt;0.001</b>	26(96.3%)	38(52.8%)	<b>&lt;0.001</b>	2(50.0%)	17(37.8%)	0.147
1-3 kg	2(11.1%)	111(48.1%)		3(9.7%)	17(16.0%)		1(7.7%)	24(23.1%)		1(3.7%)	14(19.4%)		0(0.0%)	14(31.1%)	
Does not know	1(5.6%)	21(9.1%)		0(0.0%)	12(11.3%)		0(0.0%)	6(5.8%)		0(0.0%)	5(6.9%)		0(0.0%)	9(20.0%)	
>3 kg	7(38.9%)	47(20.3%)		0(0.0%)	17(16.0%)		0(0.0%)	42(40.4%)		0(0.0%)	15(20.8%)		2(50.0%)	5(11.1%)	
<b>Mobility</b>															
Goes out	18(100%)	149(64.5%)	<b>0.005</b>	31(100%)	71(67.0%)	<b>&lt;0.001</b>	13(100%)	76(73.1%)	0.137	27(100%)	50(69.4%)	<b>0.002</b>	4(100%)	27(60.0%)	0.598
Out of bed, inside	0(0.0%)	31(13.4%)		0(0.0%)	22(20.8%)		0(0.0%)	15(14.4%)		0(0.0%)	11(15.3%)		0(0.0%)	7(15.6%)	
Bed or chair bound	0(0.0%)	51(22.1%)		0(0.0%)	13(12.3%)		0(0.0%)	13(12.5%)		0(0.0%)	11(15.3%)		0(0.0%)	11(24.4%)	
<b>Dementia/Depression</b>															
Absent	16(88.9%)	126(54.5%)	<b>0.001</b>	27(87.1%)	48(45.3%)	<b>&lt;0.001</b>	9(69.2%)	57(54.8%)	0.756	24(88.9%)	37(51.4%)	<b>0.001</b>	4(100%)	24(53.3%)	0.420
Moderate	2(11.1%)	18(7.8%)		4(12.9%)	49(46.2%)		3(23.1%)	35(33.7%)		3(11.1%)	32(44.4%)		0(0.0%)	6(13.3%)	
Severe	0(0.0%)	87(37.7%)		0(0.0%)	9(8.5%)		1(7.7%)	12(11.5%)		0(0.0%)	3(4.2%)		0(0.0%)	15(33.3%)	
<b>BMI (kg/m<sup>2</sup>)</b>															
>23	13(72.2%)	148(64.1%)	0.903	23(74.2%)	70(66.0%)	0.208	12(92.3%)	78(75.0%)	0.620	26(96.3%)	57(79.2%)	0.294	2(50.0%)	26(57.8%)	0.764
21>BMI<23	1(5.6%)	18(7.8%)		3(9.7%)	20(18.9%)		1(7.7%)	20(19.2%)		1(3.7%)	6(8.3%)		1(25.0%)	5(11.1%)	
19>BMI<21	3(16.7%)	35(15.2%)		5(16.1%)	9(8.5%)		0(0.0%)	6(5.8%)		0(0.0%)	5(6.9%)		0(0.0%)	6(13.3%)	
<19	1(5.6%)	30(13.0%)		0(0.0%)	7(6.6%)		0(0.0%)	0(0.0%)		0(0.0%)	4(5.6%)		1(25.0%)	8(17.8%)	
<b>Prescription drugs (&gt;3)</b>															
No	8(44.4%)	50(21.6%)	<b>0.028</b>	12(38.7%)	25(23.6%)	0.095	7(53.8%)	22(21.2%)	<b>0.010</b>	15(55.6%)	17(23.6%)	<b>0.002</b>	3(75.0%)	7(15.6%)	<b>0.023</b>
Yes	10(55.6%)	181(78.4%)		19(61.3%)	81(76.4%)		6(46.2%)	82(78.8%)		12(44.4%)	55(76.4%)		1(25.0%)	38(84.4%)	

**Self-rated health status**

Better	16(88.9%)	61(26.4%)	<b>&lt;0.0001</b>	12(38.7%)	31(29.2%)	<b>0.026</b>	6(46.2%)	32(30.8%)	0.170	18(66.7%)	27(37.5%)	<b>0.025</b>	3(75.0%)	10(22.2%)	0.173
As good	0(0.0%)	30(13.0%)		15(48.4%)	32(30.2%)		6(46.2%)	31(29.8%)		5(18.5%)	18(25.0%)		0(0.0%)	8(17.8%)	
Does not know	0(0.0%)	52(22.5%)		3(9.7%)	36(34.0%)		1(7.7%)	23(22.1%)		4(14.8%)	15(20.8%)		0(0.0%)	13(28.9%)	
Not as good	2(11.1%)	88(38.1%)		1(3.2%)	7(6.6%)		0(0.0%)	18(17.3%)		0(0.0%)	12(16.7%)		1(25.0%)	14(31.1%)	

**Age**

<80 years	10(55.6%)	99(42.9%)	0.421	19(61.3%)	44(41.5%)	<b>0.007</b>	3(23.1%)	49(47.1%)	0.244	21(77.8%)	48(66.7%)	0.250	2(50.0%)	17(37.8%)	1.000
80-85 years	4(22.2%)	43(18.6%)		12(38.7%)	41(38.7%)		7(53.8%)	41(39.4%)		5(18.5%)	12(16.7%)		0(0.0%)	9(20.0%)	
>85 years	4(22.2%)	89(38.5%)		0(0.0%)	21(19.8%)		3(23.1%)	14(13.5%)		1(3.7%)	12(16.7%)		2(50.0%)	19(42.2%)	

BMI, body mass index

<sup>a</sup> ≥1 impaired score in ADL, IADL, MNA, MMSE, Mini-GDS, CIRS-G, and/or TUG.

<sup>b</sup> P values from Chi2 or Fisher exact test as appropriate.

**Table A4.2.** Univariate analysis of the association between new candidate items and an abnormal Geriatric Assessment, by cancer site.

New candidate items	Digestive (n=249)			Breast (n=137)			Urinary tract (n=118)			Prostate (n=99)			Hematological (n=49)		
	Geriatric assessment			Geriatric assessment			Geriatric assessment			Geriatric assessment			Geriatric assessment		
	Normal	Abnormal <sup>a</sup>	P-value <sup>b</sup>	Normal	Abnormal <sup>a</sup>	P-value <sup>b</sup>	Normal	Abnormal <sup>a</sup>	P-value <sup>b</sup>	Normal	Abnormal <sup>a</sup>	P-value <sup>b</sup>	Normal	Abnormal <sup>a</sup>	P-value <sup>b</sup>
<b>Living alone at home</b>															
No	9(50.0%)	146(63.2%)	0.315	13(41.9%)	56(52.8%)	0.286	2(15.4%)	78(75.7%)	<b>&lt;0.0001</b>	19(70.4%)	55(76.4%)	0.539	1(25.0%)	25(55.6%)	0.330
Yes	9(50.0%)	85(36.8%)		18(58.1%)	50(47.2%)		11(84.6%)	25(24.3%)		8(29.6%)	17(23.6%)		3(75.0%)	20(44.4%)	
<b>Asthenia</b>															
No	8(44.4%)	46(19.9%)	<b>0.015</b>	16(51.6%)	34(32.1%)	<b>0.047</b>	7(53.8%)	17(16.5%)	<b>0.002</b>	16(59.3%)	22(30.6%)	<b>0.009</b>	3(75.0%)	11(25.0%)	0.069
Yes	10(55.6%)	185(80.1%)		15(48.4%)	72(67.9%)		6(46.2%)	86(83.5%)		11(40.7%)	50(69.4%)		1(25.0%)	33(75.0%)	
<b>Risk of fall<sup>c</sup></b>															
No	16(88.9%)	67(29.5%)	<b>&lt;0.0001</b>	24(77.4%)	34(32.4%)	<b>&lt;0.0001</b>	6(46.2%)	31(31.0%)	0.273	20(80.0%)	28(38.9%)	<b>&lt;0.001</b>	1(25.0%)	15(33.3%)	1.000
Yes	2(11.1%)	160(70.5%)		7(22.6%)	71(67.6%)		7(53.8%)	69(69.0%)		5(20.0%)	44(61.1%)		3(75.0%)	30(66.7%)	
<b>Fall(s) in the 6 past months</b>															
No	18(100%)	154(68.4%)	<b>0.002</b>	23(74.2%)	70(68.0%)	0.509	11(84.6%)	72(69.9%)	0.344	24(88.9%)	48(66.7%)	<b>0.041</b>	3(75.0%)	27(65.9%)	1.000
Yes	0(0.0%)	71(31.6%)		8(25.8%)	33(32.0%)		2(15.4%)	31(30.1%)		3(11.1%)	24(33.3%)		1(25.0%)	14(34.1%)	
<b>ECOG-PS</b>															
0: Fully active	10(55.6%)	29(12.6%)	<b>&lt;0.0001</b>	24(77.4%)	26(24.5%)	<b>&lt;0.0001</b>	10(76.9%)	16(15.4%)	<b>&lt;0.0001</b>	22(81.5%)	18(25.0%)	<b>&lt;0.0001</b>	4(100%)	9(20.0%)	<b>0.020</b>
1: Restricted activity	8(44.4%)	79(34.3%)		7(22.6%)	29(27.4%)		3(23.1%)	40(38.5%)		5(18.5%)	19(26.4%)		0(0.0%)	8(17.8%)	
2: Up >50% of waking hours	0(0.0%)	48(20.9%)		0(0.0%)	22(20.8%)		0(0.0%)	24(23.1%)		0(0.0%)	8(11.1%)		0(0.0%)	9(20.0%)	
3: In bed >50% of the day	0(0.0%)	58(25.2%)		0(0.0%)	21(19.8%)		0(0.0%)	14(13.5%)		0(0.0%)	19(26.4%)		0(0.0%)	9(20.0%)	
4: Completely disabled	0(0.0%)	16(7.0%)		0(0.0%)	8(7.5%)		0(0.0%)	10(9.6%)		0(0.0%)	8(11.1%)		0(0.0%)	10(22.2%)	
<b>Metastasis</b>															
No	7(41.2%)	97(45.3%)	0.741	21(77.8%)	62(66.0%)	0.243	7(53.8%)	53(58.2%)	0.764	19(73.1%)	31(47.0%)	<b>0.024</b>			
Yes	10(58.8%)	117(54.7%)		6(22.2%)	32(34.0%)		6(46.2%)	38(41.8%)		7(26.9%)	35(53.0%)				
<b>Incontinence<sup>d</sup></b>															
No	17(94.4%)	184(80.3%)	0.209	31(100%)	79(74.5%)	<b>&lt;0.001</b>	12(92.3%)	69(66.3%)	0.063	24(88.9%)	47(65.3%)	<b>0.024</b>	4(100%)	37(82.2%)	1.000
Yes	1(5.6%)	45(19.7%)		0(0.0%)	27(25.5%)		1(7.7%)	35(33.7%)		3(11.1%)	25(34.7%)		0(0.0%)	8(17.8%)	
<b>Heart failure and/or CHD</b>															
No	16(88.9%)	155(67.1%)	0.065	30(96.8%)	82(77.4%)	<b>0.015</b>	13(100%)	63(60.6%)	<b>0.004</b>	25(92.6%)	46(63.9%)	<b>0.005</b>	4(100%)	33(73.3%)	0.560
Yes	2(11.1%)	76(32.9%)		1(3.2%)	24(22.6%)		0(0.0%)	41(39.4%)		2(7.4%)	26(36.1%)		0(0.0%)	12(26.7%)	

<b>CAAF</b>															
No	15(83.3%)	181(78.7%)	0.772	29(93.5%)	87(82.1%)	0.160	13(100%)	87(84.5%)	0.210	26(96.3%)	62(87.3%)	0.276	4(100%)	28(62.2%)	0.284
Yes	3(16.7%)	49(21.3%)		2(6.5%)	19(17.9%)		0(0.0%)	16(15.5%)		1(3.7%)	9(12.7%)		0(0.0%)	17(37.8%)	
<b>Hypertension</b>															
No	10(55.6%)	75(32.5%)	<b>0.047</b>	14(45.2%)	30(28.3%)	0.077	3(23.1%)	27(26.0%)	1.000	11(40.7%)	28(38.9%)	0.867	4(100%)	17(37.8%)	<b>0.030</b>
Yes	8(44.4%)	156(67.5%)		17(54.8%)	76(71.7%)		10(76.9%)	77(74.0%)		16(59.3%)	44(61.1%)		0(0.0%)	28(62.2%)	
<b>Diabetes</b>															
No	16(88.9%)	163(70.9%)	0.169	29(93.5%)	84(79.2%)	0.104	10(76.9%)	78(75.0%)	1.000	24(88.9%)	50(69.4%)	0.068	4(100%)	36(80.0%)	1.000
Yes	2(11.1%)	67(29.1%)		2(6.5%)	22(20.8%)		3(23.1%)	26(25.0%)		3(11.1%)	22(30.6%)		0(0.0%)	9(20.0%)	
<b>Chronic renal failure</b>															
No	8(50.0%)	90(40.2%)	0.440	14(50.0%)	34(36.2%)	0.189	6(46.2%)	26(26.3%)	0.136	19(79.2%)	33(47.1%)	<b>0.006</b>	2(50.0%)	17(37.8%)	0.636
Yes	8(50.0%)	134(59.8%)		14(50.0%)	60(63.8%)		7(53.8%)	73(73.7%)		5(20.8%)	37(52.9%)		2(50.0%)	28(62.2%)	
<b>Chronic respiratory failure</b>															
No	17(94.4%)	209(91.3%)	1.000	31(100%)	101(95.3%)	0.588	13(100%)	94(90.4%)	0.600	27(100%)	70(97.2%)	1.000	4(100%)	41(91.1%)	1.000
Yes	1(5.6%)	20(8.7%)		0(0.0%)	5(4.7%)		0(0.0%)	10(9.6%)		0(0.0%)	2(2.8%)		0(0.0%)	4(8.9%)	

Abbreviations: ECOG-PS, Eastern Cooperative Oncology Group performance status; CHD, Coronary Heart Disease; CAAF, Complete Arrhythmia with Atrial Fibrillation.

<sup>a</sup>  $\geq 1$  impaired score in ADL, IADL, MNA, MMSE, Mini-GDS, CIRS-G, and/or TUG.

<sup>b</sup> P values from Chi2 or Fisher exact test as appropriate.

<sup>c</sup> Single-leg stance <5s.

<sup>d</sup> Urinary and/or fecal incontinen



## Annex 5. Logistic regression analysis of candidate items predicting impairment of the geriatric assessment.

**Table A5.1.** Logistic regression of items to predict impairment of the GA: crude odds ratios, 95%CI and p-value

	Normal GA N=97	Abnormal GA <sup>a</sup> N=632	Odds Ratio	(95% CI)	P-value
<b>ORIGINAL G8 ITEMS</b>					
<b>Anorexia</b>					<b>&lt;0.0001</b>
0: Severe	1	104	1		
1: Moderate	8	232	0.39	(0.07-2.26)	0.295
2: Absent	88	296	0.05	(0.01-0.25)	<0.001
<b>Weight loss</b>					<b>&lt;0.0001</b>
0: >3kg	2	227	1		
1: Does not know	1	63	0.47	(0.06-3.60)	0.433
2: 1-3 kg	14	119	0.09	(0.02-0.35)	0.001
3: Absent	80	223	0.03	(0.01-0.11)	<0.001
<b>BMI</b>					0.063
0: <19	3	43	1		
1: 19>BMI<21	8	67	0.64	(0.17-2.35)	0.500
2: 21>BMI<23	7	97	1.05	(0.28-3.90)	0.947
3: >23	79	425	0.43	(0.14-1.31)	0.139
<b>Mobility</b>					<b>&lt;0.001</b>
0: Bed or chair bound	0	92	1		
1: Gets out of bed/chair	0	131	1.42	(0.03-72.3)	0.861
2: Goes out	97	409	0.02	(0.001-0.37)	0.008
<b>Dementia/Depression</b>					<b>&lt;0.0001</b>
0: Severe	3	62	1		
1: Moderate	10	245	1.31	0.38-4.53)	0.801
2: Absent	84	325	0.22	(0.07-0.65)	0.006
<b>Prescription drugs</b>					
0: >3	52	499	1		
1: ≤3	45	133	0.31	(0.20-0.48)	<b>&lt;0.001</b>
<b>Self-rated health status</b>					<b>&lt;0.0001</b>
0: Not as good	1	84	1		
0.5: Does not know	8	167	0.35	(0.06-2.02)	0.241
1: As good	30	207	0.12	(0.02-0.63)	0.012
2: Better	58	174	0.05	(0.01-0.27)	<0.001
<b>Age</b>					<b>0.024</b>
0: >85 years	9	116	1		
1: 80-85 years	30	227	0.59	(0.27-1.28)	0.180
2: <80 years	58	289	0.39	(0.19-0.81)	0.110
<b>NEW CANDIDATE ITEMS</b>					
Not living alone	47	407	1		
Living alone at home	50	224	0.52	(0.34-0.80)	<b>0.002</b>
No asthenia	52	143	1		
Asthenia	45	487	3.94	(2.53-6.11)	<b>&lt;0.001</b>
No risk of fall	69	188	1		
Risk of fall <sup>b</sup>	26	435	6.14	(3.79-9.94)	<b>&lt;0.001</b>
No falls	83	407	1		
Fall(s) in the 6 past months	14	209	3.04	(1.69-5.50)	<b>&lt;0.001</b>

<b>ECOG-PS</b>					<b>&lt;0.0001</b>
0: Fully active	71	106	1		
1: Restricted activity	24	189	5.19	(3.1-8.7)	<0.001
2: Up >50% of waking hours	2	127	34.24	(9.5-124)	<0.001
3: In bed >50% of the day	0	148	199.39	(12.2-3254)	<0.001
4: Completely disabled	0	61	82.58	(5.0-1356)	0.002
No metastases	55	264	1		
Metastases	32	267	1.74	(1.09-2.78)	<b>0.020</b>
Not incontinent	91	471	1		
Incontinence	6	159	5.12	(2.20-11.92)	<b>&lt;0.001</b>
No heart failure / CHD	91	434	1		
Heart failure / CHD	6	198	6.92	(2.98-16.08)	<b>&lt;0.001</b>
No CAAF	91	503	1		
CAAF	6	126	3.8	(1.63-8.88)	<b>0.002</b>
No Heart rhythm disorder	76	527	1		
Heart rhythm disorder	21	102	0.69	(0.41-1.17)	0.167
No hypertension	44	201	1		
Hypertension	53	431	1.78	(1.15-2.75)	<b>0.008</b>
No diabetes	87	471	1		
Diabetes	10	160	2.96	(1.50-5.83)	<b>0.002</b>
No renal failure	51	229	1		
Chronic renal failure	37	376	2.26	(1.44-3.56)	<b>&lt;0.001</b>
No respiratory failure	96	580	1		
Chronic respiratory failure	1	49	5.49	(1.07-28.24)	<b>0.042</b>

Abbreviations: GA, geriatric assessment; 95% CI, 95% confidence interval; BMI, body mass index; ECOG-PS, Eastern Cooperative Oncology Group performance status; CHD, coronary heart disease; CAAF, complete arrhythmia with atrial fibrillation.

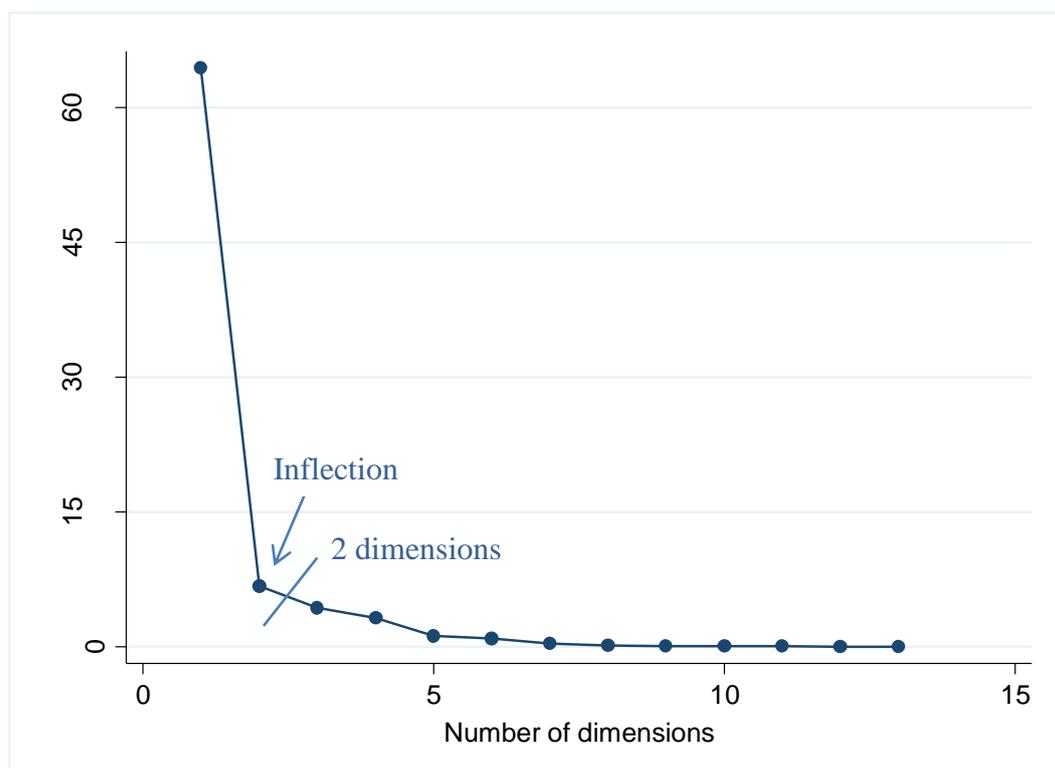
<sup>a</sup> ≥1 impaired score in ADL, IADL, MNA, MMSE, Mini-GDS, CIRS-G, and/or TUG.

<sup>b</sup> Single-leg stance <5s.

## **Annex 6. Multiple Correspondence Analysis**

### Choice of the number of dimensions to interpret

As a first step, MCA was applied to the data and the number of dimensions useful for interpretation was retained. Commonly used rules recommend that the number of dimensions retained represent >70% of the inertia<sup>1</sup> or correspond to the elbow criterion (number right before the inflection point, followed by a steady decline in a scree plot of inertias). **Figure A6.1** shows the Scree plot presenting the proportions of inertia (variance) explained by each dimension. Two dimensions are suggested, with 72% of the total inertia explained by the 2 first factors corresponding to the number before the inflection point.



**Figure A6.1.** Scree plot of principal inertias after MCA.

**Table A6.1**, and **Figures A6.2** and **A6.3** show the results from the MCA.

<sup>1</sup> Higgs NT. Practical and innovative uses of correspondence analysis. *The Statistician*. 1991;40(2):183-94.

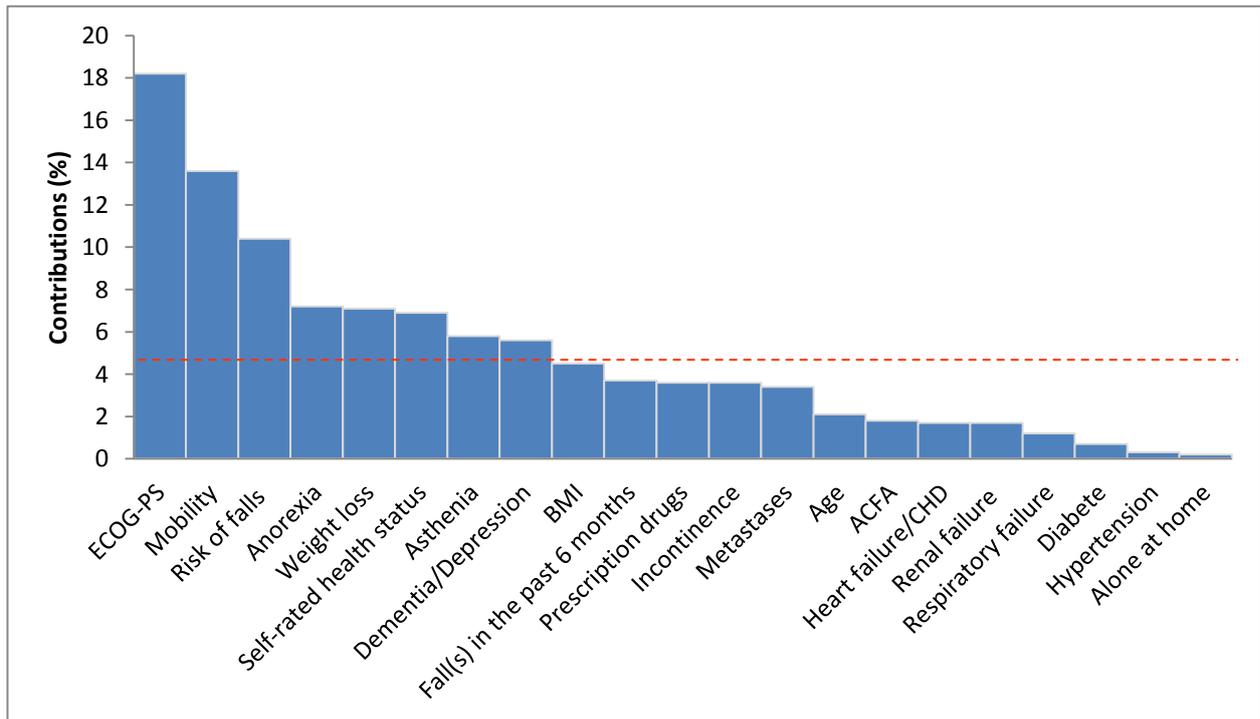
**Table A6.1.** Quality and contribution of categories in the composition of the first two dimensions from multiple correspondence analysis.

Items	Dimension 1		Dimension 2		
	Sqcorr <sup>a</sup>	Contribution <sup>b</sup>	Sqcorr <sup>a</sup>	Contribution <sup>b</sup>	
Anorexia	Severe	<b>0.539</b>	<b>0.024</b>	0.062	0.017
	Moderate	<b>0.731</b>	0.015	0.000	0.000
	Absent	<b>0.759</b>	<b>0.033</b>	0.018	0.007
Weight loss	>3kg	<b>0.673</b>	<b>0.026</b>	0.017	0.006
	Does not know	<b>0.627</b>	0.009	0.062	0.009
	1-3kg	0.039	0.000	0.034	0.002
	Absent	<b>0.721</b>	<b>0.036</b>	0.039	0.017
Body Mass Index	<19	0.405	0.017	0.231	0.007
	19-21	0.086	0.011	0.340	0.001
	21-23	0.243	0.010	0.348	0.002
	>23	0.300	0.016	0.403	0.003
Mobility	Bed/chair bound	<b>0.694</b>	<b>0.067</b>	0.000	0.000
	Gets out of bed/chair	<b>0.684</b>	<b>0.031</b>	0.032	0.014
	Goes out	<b>0.851</b>	<b>0.038</b>	0.008	0.003
ECOG-PS	0: Fully active	<b>0.804</b>	<b>0.069</b>	0.000	0.000
	1: Restricted activity	<b>0.538</b>	0.013	0.041	0.009
	2: Up >50% of the day	0.401	0.009	0.010	0.002
	3: In bed >50% / day	<b>0.781</b>	<b>0.049</b>	0.005	0.003
Asthenia	4: Completely disabled	<b>0.583</b>	<b>0.042</b>	0.000	0.000
	No	<b>0.828</b>	<b>0.042</b>	0.001	0.000
Fall(s) in the past 6 months	Yes	<b>0.828</b>	0.016	0.001	0.000
	No	<b>0.851</b>	0.012	0.001	0.000
Risk of falls <sup>c</sup>	Yes	<b>0.851</b>	<b>0.025</b>	0.001	0.000
	No	<b>0.853</b>	<b>0.066</b>	0.004	0.003
Dementia/Depression	Yes	<b>0.853</b>	<b>0.038</b>	0.004	0.001
	Severe	<b>0.692</b>	0.014	0.003	0.001
	Moderate	<b>0.821</b>	<b>0.019</b>	0.000	0.000
Self-rated health status	Absent	<b>0.875</b>	<b>0.023</b>	0.001	0.000
	Not as good	<b>0.633</b>	<b>0.021</b>	0.083	0.016
	Does not know	<b>0.752</b>	<b>0.019</b>	0.002	0.001
	As good	0.260	0.002	0.157	0.010
Prescription drugs	Better	<b>0.853</b>	<b>0.027</b>	0.001	0.000
	<6	0.234	0.018	<b>0.582</b>	<b>0.079</b>
Age	≥6	0.234	0.018	<b>0.582</b>	<b>0.068</b>
	>85 years	0.317	0.007	0.007	0.001
	80-85 years	0.414	0.004	0.094	0.018
	<80 years	0.453	0.010	0.186	<b>0.019</b>
Heart failure/coronary heart disease	No	0.388	0.005	<b>0.528</b>	<b>0.049</b>
	Yes	0.388	0.012	<b>0.528</b>	<b>0.128</b>
Hypertension	No	0.114	0.002	<b>0.551</b>	<b>0.102</b>
	Yes	0.114	0.001	<b>0.551</b>	<b>0.051</b>
Diabete	No	0.249	0.002	0.422	<b>0.026</b>
	Yes	0.249	0.005	0.422	<b>0.085</b>
Respiratory failure	No	0.191	0.001	<b>0.545</b>	0.002
	Yes	0.191	0.011	<b>0.545</b>	<b>0.036</b>
Renal failure	No	0.468	0.010	0.052	<b>0.019</b>
	Yes	0.468	0.007	0.052	0.013

Sqcorr, squared correlation; ECOG-PS, Eastern Cooperative Oncology Group performance status.

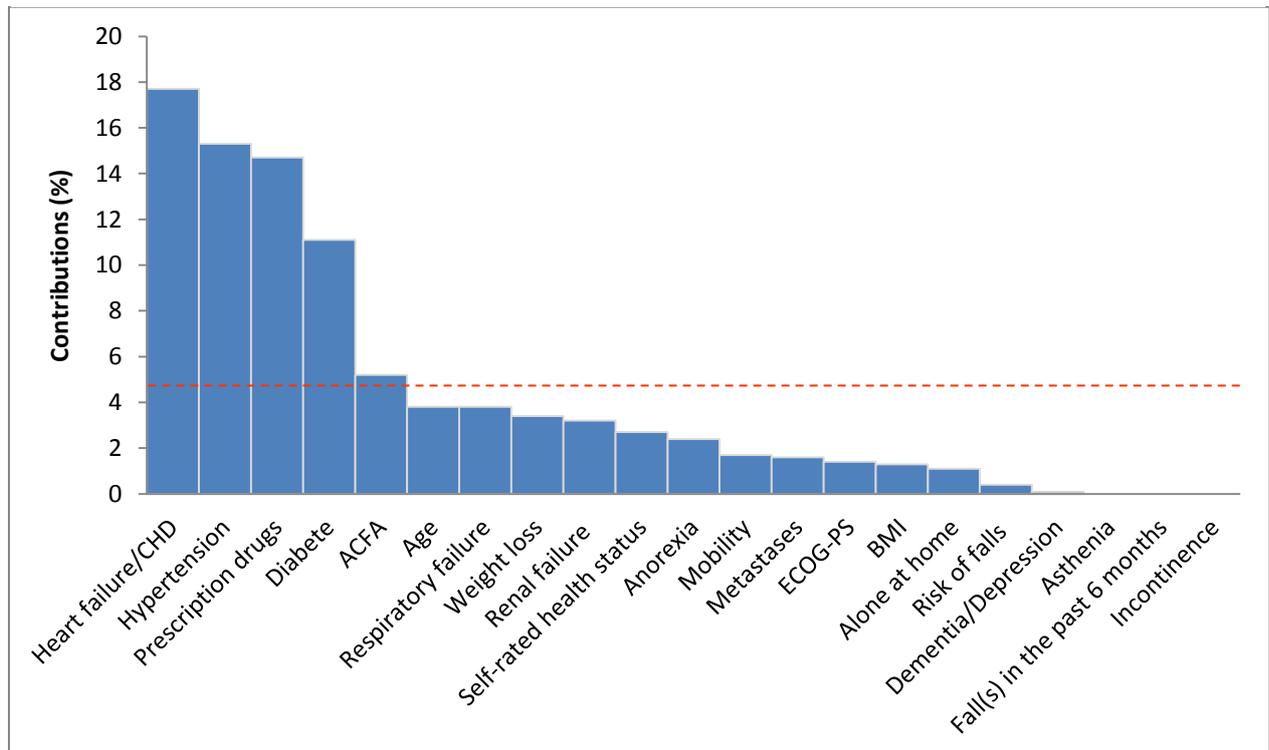
<sup>a</sup> Quality of representation of factors. Comprised between 0 and 1, the highest correlations (>0.5) are marked in bold.

<sup>b</sup> Contributions above the average (>0.018) are marked in bold.



ECOG-PS, Eastern Cooperative Oncology Group-performance status; BMI, Body Mass Index; ACFA, Complete arrhythmia with atrial fibrillation; CHD, Coronary Heart Disease.

Figure A6.2. Contributions of variables to the Dimension 1.



ECOG-PS, Eastern Cooperative Oncology Group-performance status; BMI, Body Mass Index; ACFA, Complete arrhythmia with atrial fibrillation; CHD, Coronary Heart Disease

Figure A6.3. Contribution of variables to the Dimension 2.

**Note:** the red dashed lines on the graphs above indicate the expected average contribution according to the total number of variables. Categories with a contribution larger than this threshold could be considered as important in contributing to that dimension.

## **Annex 7. Complementary results from Gold Standard analysis**

**Table A7. 1.** Associations between indicators and latent class profiles

		Latent typology				<i>p</i> -values
		Fit (LC1) ( <i>N</i> =233)		Unhealthy profiles (LC2-LC4) ( <i>N</i> =903)		
<b>Indicators and covariates</b>		<i>N</i>	%	<i>N</i>	%	<i>p</i> <i>Chi</i> <sup>2</sup>
<b>Functional impairment</b>	ADL≤5/6	10	4.3	349	38.6	<b>&lt;0.0001</b>
<b>Cognitive impairment</b>	MMSE≤23/30	32	13.7	288	31.9	<b>&lt;0.0001</b>
<b>Malnutrition</b>	HAS criteria	0	0.0	776	85.9	<b>&lt;0.0001</b>
<b>Social environment</b>	Inadequate	18	7.7	159	17.6	<b>0.0002</b>
<b>Depression</b>	DSMIV criteria	26	11.2	338	37.4	<b>&lt;0.0001</b>
<b>No of severe comorbidities (grade 3-4 CIRS-G)</b>						
	0	176	75.5	237	26.2	<b>&lt;0.0001</b>
	1	52	22.3	251	27.8	
	≥2	5	2.1	415	46.0	
Tumour site						
	Colorectal	20	8.6	181	20.0	<b>&lt;0.0001</b>
	Breast	82	35.2	100	11.1	
	Prostate	62	26.6	65	7.2	
	Gastrointestinal	13	5.6	199	22.0	
	Urinary	33	14.2	134	14.8	
	Hematological	2	0.9	82	9.1	
	Others	21	9.0	142	15.7	
Metastatic status						
	M0	151	64.8	277	30.7	<b>&lt;0.0001</b>
	M1/Mx	58	24.9	436	48.3	
	Not reported or NA	24	10.3	190	21.0	
Male Gender		112	48.1	475	52.6	0.217
Age						
	≤ 80 years	136	58.4	443	49.1	<b>0.011</b>
	> 80 years	97	41.6	460	50.9	
In/Outpatient status at the time of GA						
	Outpatient	233	100.0	491	54.4	<b>&lt;0.0001</b>
	Inpatient	0	0.0	412	45.6	

Abbreviations: ADL, Activities of Daily Living; MMSE, Mini-Mental State Evaluation; HAS, French National Authority for Health; DSM, Diagnostic and Statistical Manual of Mental Disorders; CIRS-G, Cumulative Illness Rating Scale for Geriatrics; M0, absence of distant metastases; M1, presence of distant metastases; Mx, metastatic status not assessable; NA, not applicable; GA, Geriatric Assessment.

**Table A7.2.** Diagnostic performances according to different reference standards (original vs. modified G8) with available data on the definition tested and both scores.

Reference standards	Prevalence	Population n	Screening tools	Cut-offs <sup>a</sup>	Sensitivity (95% CI)	Specificity (95% CI)	AUROC (95% CI)	AUROC p-values
<b>GA <math>\geq 1</math> impairment</b>	92.4%	1625	Original G8	$\leq 14$	89.6% (88.0%-91.1%)	57.3% (48.1%-66.1%)	0.89 (0.87-0.91)	<b>0.0003</b>
			Modified G8	$\geq 6$	90.8% (89.2%-92.2%)	75.0% (66.4%-82.3%)	0.93 (0.91-0.94)	
<b>GA <math>\geq 2</math> impairments</b>	76.5%	1624	Original G8	$\leq 13.5$	88.2% (86.2%-89.9%)	57.2% (52.1%-62.2%)	0.86 (0.84-0.88)	<b>0.0001</b>
			Modified G8	$\geq 8$	92.0% (90.4%-93.5%)	61.9% (56.9%-66.8%)	0.89 (0.88-0.91)	
<b><math>\geq 1</math> geriatric intervention</b>	88.4%	1673	Original G8	$\leq 14$	89.0% (87.3%-90.6%)	46.9% (39.7%-54.2%)	0.82 (0.79-0.85)	<b>0.0158</b>
			Modified G8	$\geq 6$	89.2% (87.5%-90.7%)	52.1% (44.8%-59.3%)	0.84 (0.82-0.87)	
<b>SIOG classification <sup>b</sup></b>	83.7%	1541	Original G8	$\leq 13.5$	84.3% (82.2%-86.3%)	59.9% (53.6%-66.0%)	0.84 (0.81-0.86)	<b>&lt;0.00001</b>
			Modified G8	$\geq 8$	87.6% (85.7%-89.3%)	68.3% (62.1%-74.0%)	0.88 (0.86-0.90)	
<b>Balducci's classification <sup>c</sup></b>	86.2%	1632	Original G8	$\leq 13.5$	82.6% (80.5%-84.5%)	60.6% (53.9%-67.0%)	0.81 (0.79-0.84)	0.1700
			Modified G8	$\geq 8$	84.6% (82.6%-86.5%)	60.6% (53.9%-67.0%)	0.83 (0.80-0.86)	
<b>Latent class typology <sup>d</sup></b>	79.0%	1151	Original G8	$\leq 13.5$	87.8% (85.5%-89.8%)	63.6% (57.2%-69.7%)	0.86 (0.84-0.88)	0.1654
			Modified G8	$\geq 8$	88.0% (85.7%-90.0%)	58.7% (52.2%-64.9%)	0.85 (0.82-0.87)	

Abbreviations: AUROC, area under the ROC (receiver operating characteristic) curve; 95% CI, 95% confidence interval; GA, Geriatric Assessment; SIOG, International Society of Geriatric Oncology

<sup>a</sup> Official cut-off for GA  $\geq 1$  impairment and best cut-off (prioritizing sensitivity) otherwise.

<sup>b</sup> Fit vs. Unfit (Vulnerable or frail or too sick)

<sup>c</sup> Fit vs. Unfit (Vulnerable or frail)

<sup>d</sup> Fit vs. Unfit (Latent classes 2 to 4: Malnourished or cognitive/mood impaired or globally impaired)



## **Annex 8. Complementary results from Survival analysis**

**Table A8.1.** Prognostic value of G8: results according to study population with respect to G8/ modified G8 availability.

Outcome		Unadjusted analysis						Adjusted analysis					
		Patients with available G8 and mG8			Patients with available G8			Patients with available G8 and mG8			Patients with available G8		
		HR (95% CI)	Harrell-C	Gönen <sup>a</sup>	HR (95% CI)	Harrell-C	Gönen <sup>a</sup>	aHR (95% CI)	Harrell-C	Gönen <sup>a</sup>	aHR (95% CI)	Harrell-C	Gönen <sup>a</sup>
<i>1-year survival</i>	Abnormal ≤14	5.7 (3.8-8.5)	0.58	0.60	5.3 (3.6-7.7)	0.58	0.59	4.3 (2.7-6.8)	0.74	0.73	3.9 (2.5-6.0)	0.74	0.73
	13.5-17	1(ref)	0.71	0.70	1(ref)	0.71	0.70	1(ref)	0.79	0.77	1(ref)	0.78	0.76
	10-13	3.1 (2.2-4.2)			2.9 (2.1-3.9)			2.7 (1.9-3.9)			2.7 (1.9-3.7)		
	8-9.5	6.0 (4.3-8.4)			5.6 (4.1-7.7)			4.5 (3.1-6.5)			4.5 (3.2-6.4)		
	6-7.5	8.4 (5.9-11.9)			8.2 (5.9-11.4)			5.7 (3.8-8.4)			6.0 (4.1-8.7)		
	0-5.5	14.8 (10.3-21.0)			14.0 (10.0-19.6)			10.3 (6.8-15.4)			10.0 (6.5-14.0)		
<i>3-years survival</i>	Abnormal ≤14	4.0 (3.1-5.3)	0.58	0.58	3.9 (3.0-5.1)	0.58	0.58	2.9 (2.2-4.0)	0.75	0.73	3.0 (2.2-4.0)	0.73	0.72
	13.5-17	1(ref)	0.69	0.67	1(ref)	0.69	0.67	1(ref)	0.77	0.74	1(ref)	0.76	0.74
	10-13	2.3 (1.8-2.9)			2.2 (1.8-2.8)			1.9 (1.5-2.5)			1.9 (1.5-2.5)		
	8-9.5	3.7 (2.9-4.7)			3.7 (2.9-4.6)			2.7 (2.1-3.5)			2.8 (2.2-3.6)		
	6-7.5	5.5 (4.2-7.1)			5.5 (4.3-7.1)			3.8 (2.8-5.1)			4.0 (3.0-5.3)		
	0-5.5	9.3 (7.1-12.3)			9.0 (6.9-11.7)			6.8 (5.0-9.4)			6.5 (4.8-8.8)		

Abbreviations: 95% CI, 95% confidence interval; mG8, modified G8; HR, unadjusted hazard ratio; aHR, hazard ratio adjusted for age, gender, anticancer treatment, cancer site and metastasis;

<sup>a</sup> Gönen and Heller's K concordance statistic

**Table A8.2.** Prognostic values for overall 1- and 3-years mortality of the ECOG-PS and the TUG: Cox proportional hazards models

Outcome	Screening tools	Unadjusted analysis N=1333								Adjusted analysis N=1168					
		N (%)	Deaths (%)	HR	95% CI	p-value	Harrell-C	Gönen <sup>a</sup>	N (%)	aHR	95% CI	p-value	Harrell-C	Gönen <sup>a</sup>	
1-year survival	ECOG-PS	0	276 (21)	40 (14)	1(ref)		<0.0001	0.72	0.69	247 (21)	1(ref)		<0.0001	0.78	0.76
		1	383 (29)	107 (28)	2.1	1.5-3.0				335 (29)	2.1	1.5-3.2			
		2	240 (18)	96 (40)	3.3	2.3-4.8				207 (18)	2.8	1.9-4.2			
		3	295 (22)	192 (65)	7.5	5.4-10.6				263 (22)	5.9	4.1-8.4			
		4	139 (10)	102 (73)	11.0	7.6-15.9				116 (10)	8.9	6.0-13.3			
	TUG	Abnormal ≥20	519 (39)	288 (55)	2.3	1.9-2.7	<0.0001	0.60	0.59	445 (38)	1.8	1.5-2.2	<0.0001	0.74	0.73
3-years survival	ECOG-PS	0	276 (21)	95 (34)	1(ref)		<0.0001	0.69	0.66	247 (21)	1(ref)		<0.0001	0.76	0.74
		1	383 (29)	179 (47)	1.6	1.2-2.0				335 (29)	1.5	1.1-1.9			
		2	240 (18)	144 (60)	2.4	1.8-3.1				207 (18)	1.9	1.4-2.5			
		3	295 (22)	220 (75)	4.5	3.6-5.8				263 (22)	3.5	2.6-4.5			
		4	139 (10)	121 (87)	7.2	5.5-9.4				116 (10)	6.0	4.4-8.3			
	TUG	Abnormal ≥20	519 (39)	357 (69)	1.9	1.7-2.2	<0.0001	0.59	0.58	445 (38)	1.6	1.4-1.9	<0.0001	0.72	0.71

Abbreviations: ECOG-PS, Eastern Cooperative Oncology Group Performance Status; TUG, timed up-and-go test

HR, unadjusted hazard ratio; aHR, hazard ratio adjusted for age, gender, anticancer treatment, cancer site and metastasis; 95% CI, 95% confidence interval.

<sup>a</sup> Gönen and Heller's K concordance statistic

**Table A8.3.** Prognostic value for overall 1- and 3-years survival by tumor site, using scores (G8 and modified G8) in classes of increasing risk: adjusted Cox proportional-hazards models

Cancer site	Colorectal			Upper digestive tract and liver			Urinary tract			Prostate			Hematological malignances			Breast		
	N (%D)	aHR (95% CI)	p-value	N (%D)	aHR (95% CI)	p-value	N (%D)	aHR (95% CI)	p-value	N (%D)	aHR (95% CI)	p-value	N (%D)	aHR (95% CI)	p-value	N (%D)	HRa (IC95%)	p-value
<b>1-year Survival G8</b>																		
13.5-17	47 (15)	1(ref)	<b>&lt;0.001</b>	24 (38)	1(ref)	<b>&lt;0.001</b>	50 (18)	1(ref)	<b>&lt;0.001</b>	75 (5)	1(ref)	<b>0.002</b>	17 (0)	1(ref)	<b>&lt;0.001</b>	80 (3)	1(ref)	<b>0.001</b>
10-13	103 (23)	1.8 (0.77-4.4)		68 (46)	1.4 (0.64-2.8)		67 (48)	3.4 (1.6-7.3)		34 (38)	5.0 (1.51-16)		37 (35)	10 (1.36-138)		55 (20)	10 (2.2-48)	
8-9.5	46 (39)	4.2 (1.7-10)		50 (72)	2.9 (1.4-6.2)		28 (54)	3.8 (1.6-9.1)		15 (67)	7.4 (2.07-26)		22 (55)	24 (3.19-312)		26 (35)	7.2 (1.4-37)	
6-7.5	27 (59)	7.8 (3.0-20)		32 (72)	2.9 (1.3-6.5)		10 (90)	11 (4.1-30)		3 (67)	33 (5.13-215)		19 (47)	28 (3.47-358)		14 (50)	15 (3.0-81)	
0-5.5	15 (73)	4.4 (1.6-12)		17 (100)	6.7 (2.8-15)		11 (82)	16 (6.0-44)		14 (64)	9.8 (2.2-43)		10 (90)	58 (6.79-753)		11 (73)	31 (5.9-165)	
<b>Modified G8</b>																		
0-7	50 (12)	1(ref)	<b>0.008</b>	22 (36)	1(ref)	<b>&lt;0.001</b>	40 (23)	1(ref)	<b>0.002</b>	55 (4)	1(ref)	<b>0.003</b>	15 (0)	1(ref)	<b>&lt;0.001</b>	70 (0)	1(ref)	<b>&lt;0.001</b>
8-13	35 (11)	1.0 (0.28-3.6)		22 (27)	0.84 (0.28-2.5)		26 (31)	1.4 (0.55-3.8)		19 (11)	1.9 (0.26-13)		11 (36)	8.3 (0.82-112)		22 (5)	2.2 (0.20-16)	
14-20	49 (37)	3.3 (1.3-8.5)		42 (52)	1.7 (0.76-3.9)		40 (43)	2.6 (1.2-5.9)		29 (31)	8.7 (1.7-45)		21 (24)	5.4 (0.57-725)		47 (26)	7.3 (2.0-40)	
21-29	80 (45)	3.6 (1.5-8.7)		74 (70)	2.9 (1.4-6.2)		47 (64)	3.9 (1.8-8.4)		26 (62)	14 (2.8-75)		32 (53)	19 (2.5-532)		32 (47)	18 (5.1-102)	
30-35	24 (50)	4.5 (1.6-12)		31 (90)	5.4 (2.4-12)		13 (77)	5.1 (1.9-14)		12 (75)	20 (3.7-112)		26 (65)	27 (3.5-749)		15 (60)	25 (6.4-148)	
<b>3-years Survival G8</b>																		
13.5-17	47 (45)	1(ref)	<b>&lt;0.001</b>	24 (63)	1(ref)	<b>&lt;0.001</b>	50 (50)	1(ref)	<b>&lt;0.001</b>	75 (21)	1(ref)	<b>&lt;0.001</b>	17 (18)	1(ref)	<b>&lt;0.001</b>	80 (13)	1(ref)	<b>&lt;0.001</b>
10-13	103 (49)	1.2 (0.69-2.0)		68 (74)	1.4 (0.75-2.4)		67 (61)	1.9 (1.1-3.1)		34 (59)	3.2 (1.5-6.8)		37 (51)	3.1 (0.88-11)		55 (38)	4.1 (1.8-9.0)	
8-9.5	46 (59)	1.9 (1.0-3.5)		50 (82)	2.1 (1.2-3.9)		28 (64)	2.2 (1.1-4.3)		15 (80)	3.7 (1.6-8.8)		22 (73)	6.7 (1.9-23)		26 (58)	2.8 (1.2-6.9)	
6-7.5	27 (81)	4.0 (2.1-7.7)		32 (75)	2.6 (1.3-5.0)		10 (100)	7.0 (3.1-15)		3 (67)	21.3 (4.1-109)		19 (63)	7.1 (2.0-26)		14 (71)	7.8 (3.0-20)	
0-5.5	15 (93)	3.7 (1.8-7.7)		17 (100)	5.9 (2.8-12)		11 (82)	9.9 (4.2-23)		14 (79)	6.9 (2.2-21)		10 (90)	16 (4.1-65)		11 (82)	8.6 (3.1-24)	
<b>Modified G8</b>																		
0-7	50 (44)	1(ref)	<b>0.023</b>	22 (68)	1(ref)	<b>&lt;0.001</b>	40 (55)	1(ref)	<b>0.001</b>	55 (16)	1(ref)	<b>&lt;0.001</b>	15 (13)	1(ref)	<b>0.001</b>	70 (9)	1(ref)	<b>&lt;0.001</b>
8-13	35 (37)	0.94 (0.47-1.9)		22 (77)	1.2 (0.57-2.4)		26 (38)	0.76 (0.36-1.6)		19 (42)	2.7 (1.0-7.1)		11 (36)	2.0 (0.34-11)		22 (23)	2.9 (0.85-9.6)	
14-20	49 (61)	1.7 (0.94-2.9)		42 (67)	1.2 (0.63-2.3)		40 (60)	1.6 (0.91-3.0)		29 (52)	4.4 (1.8-10)		21 (57)	3.5 (0.75-16)		47 (45)	4.0 (1.5-10)	
21-29	80 (63)	1.6 (0.92-2.6)		74 (80)	2.0 (1.1-3.5)		47 (77)	2.5 (1.4-4.4)		26 (73)	5.8 (2.4-14)		32 (66)	6.6 (1.5-28)		32 (69)	7.8 (3.0-20)	
30-35	24 (79)	2.8 (1.4-5.5)		31 (90)	3.7 (1.9-7.4)		13 (85)	3.5 (1.6-7.9)		12 (83)	11 (3.9-32)		26 (77)	10 (2.4-45)		15 (73)	7.4 (2.6-21)	

Abbreviations: aHR, HR adjusted for age, gender, anticancer treatment and metastasis; 95% CI, 95% confidence interval.

**Table A8.4.** Harrell's C-index and Gönen and Heller's K concordance statistic for G8 and modified G8 by tumor site (overall 1-year survival analysis)

Screening tools		Unadjusted analysis N=1333			Adjusted analysis <sup>a</sup> N=1168		
		Tumor site	N	C-statistic	K-statistic	N	C-statistic
<b>G8</b>							
<i>Abnormal G8 model</i>	Colorectal	261	0.54	0.55	238	0.75	0.74
	Upper digestive tract and liver	228	0.52	0.51	191	0.70	0.68
	Breast	218	0.64	0.69	186	0.84	0.86
	Urinary tract	197	0.58	0.58	166	0.70	0.69
	Prostate	155	0.72	0.71	141	0.85	0.81
	Hematological	105	0.56	0.59	105	0.69	0.71
	Others	169	0.54	0.54	141	0.69	0.68
<i>5 classes<sup>b</sup> G8 score model</i>	Colorectal	261	0.68	0.66	238	0.79	0.77
	Upper digestive tract and liver	228	0.64	0.64	191	0.73	0.72
	Breast	218	0.76	0.73	186	0.87	0.84
	Urinary tract	197	0.69	0.68	166	0.75	0.73
	Prostate	155	0.81	0.75	141	0.88	0.81
	Hematological	105	0.72	0.72	105	0.75	0.78
	Others	169	0.66	0.65	141	0.73	0.71
<b>Modified G8</b>							
<i>Abnormal modified G8 model</i>	Colorectal	261	0.55	0.56	238	0.77	0.74
	Upper digestive tract and liver	228	0.52	0.52	191	0.71	0.69
	Breast	218	0.64	0.67	186	0.85	0.87
	Urinary tract	197	0.58	0.58	166	0.70	0.68
	Prostate	155	0.69	0.70	141	0.86	0.83
	Hematological	105	0.58	0.61	105	0.70	0.72
	Others	169	0.54	0.54	141	0.69	0.68
<i>5 classes<sup>c</sup> modified G8 score model</i>	Colorectal	261	0.68	0.67	238	0.79	0.76
	Upper digestive tract and liver	228	0.67	0.66	191	0.74	0.72
	Breast	218	0.79	0.76	186	0.89	0.88
	Urinary tract	197	0.69	0.67	166	0.72	0.70
	Prostate	155	0.82	0.77	141	0.89	0.83
	Hematological	105	0.70	0.72	105	0.76	0.79
	Others	169	0.68	0.66	141	0.75	0.73

C-statistic, Harrell's C-index; K-statistic, Gönen and Heller's K concordance index; 95% CI, 95% confidence interval.

<sup>a</sup> Adjusted for age, gender, metastasis status and anticancer treatment.

<sup>b</sup> 5 classes of increasing risk: 13.5-17, 10-13, 8-9.5, 6-7.5, 0-5.5.

<sup>c</sup> 5 classes of increasing risk: 0-7, 8-13, 14-20, 21-29, 30-35.

**Table A8.5.** Harrell's C-index and Gönen and Heller's K concordance statistic for G8 and modified G8 by tumor site (overall 3-years survival analysis)

Screening tools		Unadjusted analysis N=1333			Adjusted analysis <sup>a</sup> N= 1168			
		Tumor site	N	C-statistic	K-statistic	N	C-statistic	K-statistic
<b>G8</b>								
<i>Abnormal G8 model</i>	Colorectal	261	0.54	0.54	238	0.74	0.72	
	Upper digestive tract and liver	228	0.52	0.51	191	0.69	0.66	
	Breast	218	0.64	0.66	186	0.81	0.80	
	Urinary tract	197	0.57	0.55	166	0.67	0.64	
	Prostate	155	0.72	0.69	141	0.84	0.80	
	Hematological	105	0.57	0.59	105	0.68	0.70	
	Others	169	0.54	0.54	141	0.69	0.66	
	<i>5 classes<sup>b</sup> G8 score model</i>	Colorectal	261	0.65	0.62	238	0.76	0.74
		Upper digestive tract and liver	228	0.64	0.63	191	0.71	0.70
		Breast	218	0.74	0.70	186	0.84	0.79
		Urinary tract	197	0.67	0.64	166	0.71	0.69
		Prostate	155	0.77	0.73	141	0.86	0.79
		Hematological	105	0.70	0.68	105	0.72	0.73
		Others	169	0.66	0.63	141	0.72	0.70
<b>Modified G8</b>								
<i>Abnormal modified G8 model</i>	Colorectal	261	0.52	0.51	238	0.73	0.72	
	Upper digestive tract and liver	228	0.52	0.52	191	0.70	0.66	
	Breast	218	0.63	0.64	186	0.82	0.81	
	Urinary tract	197	0.56	0.54	166	0.66	0.64	
	Prostate	155	0.68	0.66	141	0.85	0.81	
	Hematological	105	0.58	0.59	105	0.68	0.69	
	Others	169	0.54	0.54	141	0.69	0.67	
	<i>5 classes<sup>c</sup> modified G8 score model</i>	Colorectal	261	0.64	0.61	238	0.75	0.73
		Upper digestive tract and liver	228	0.65	0.62	191	0.72	0.69
		Breast	218	0.75	0.73	186	0.85	0.80
		Urinary tract	197	0.66	0.64	166	0.69	0.67
		Prostate	155	0.78	0.73	141	0.86	0.81
		Hematological	105	0.68	0.67	105	0.72	0.73
		Others	169	0.68	0.65	141	0.73	0.71

C-statistic, Harrell's C-index; K-statistic, Gönen and Heller's K concordance index; 95% CI, 95% confidence interval.

<sup>a</sup> Adjusted for age, gender, metastasis status and anticancer treatment.

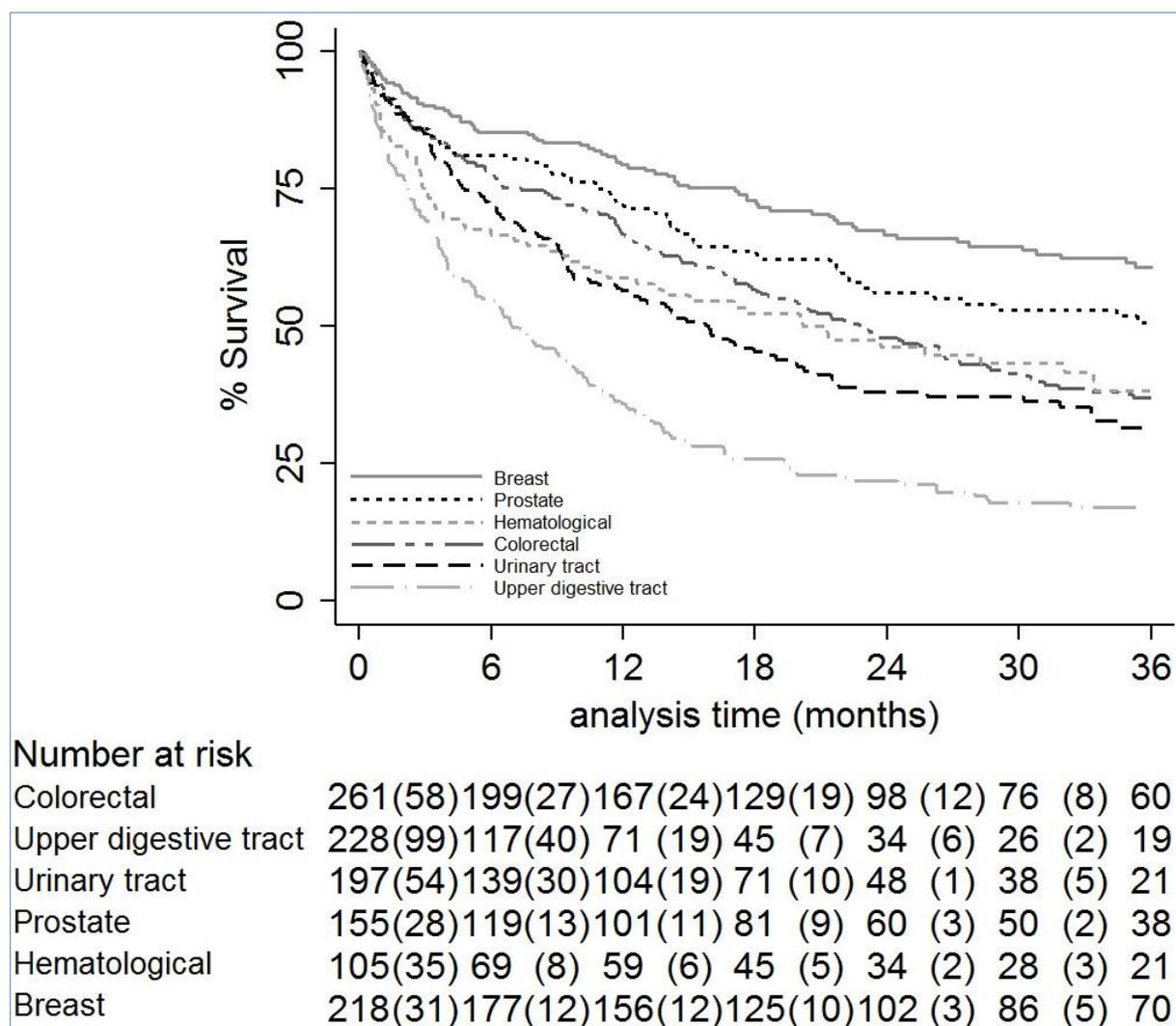
<sup>b</sup> 5 classes of increasing risk: 13.5-17, 10-13, 8-9.5, 6-7.5, 0-5.5.

<sup>c</sup> 5 classes of increasing risk: 0-7, 8-13, 14-20, 21-29, 30-35.

**Table A8.6.** Patient characteristics comparisons between patients with and without digestive cancers.

Characteristics	Non digestive cancers	Colorectal cancer	p-value	Upper digestive tract cancer	p-value	
	n=844	n=261		n=228		
	N (%)	N (%)		N (%)		
Age, median (IQR)	80 (76;84)	80 (76-84)	0.488	79 (76;83)	0.377	
Male gender	446 (52.8)	121 (46.4)	0.067	123 (53.9)	0.767	
Metastasis	299 (47.2)	128 (53.8)	0.081	97 (50.8)	0.379	
No. of drugs, median (IQR)	6 (4;8)	6 (4;8)	0.192	6 (4;9)	0.646	
No. of comorbidities, median (IQR)	2 (1;3)	2.5 (1;4)	0.098	2 (1;3)	0.803	
ECOG-PS			<b>&lt;0.0001</b>		<b>&lt;0.0001</b>	
0 - Fully active	210 (24.9)	43 (16.5)		23 (10.1)		
1 - Restricted activity but ambulatory	216 (25.6)	86 (33.0)		81 (35.5)		
2 - Up >50% of waking hours	143 (16.9)	50 (19.2)		47 (20.6)		
3 - Confined to bed >50% of the day	174 (20.6)	64 (24.5)		57 (25.0)		
4 - Completely disabled	101 (12.0)	18 (6.9)		20 (8.8)		
Abnormal Geriatric Assessment	705 (88.8)	243 (96.4)	<b>&lt;0.001</b>	222 (98.7)	<b>&lt;0.0001</b>	
ADL≤5	290 (34.5)	80 (30.7)	0.247	70 (30.7)	0.279	
IADL≤7	480 (61.5)	149 (68.3)	0.066	134 (66.7)	0.180	
MMSE≤23	171 (27.6)	37 (18.6)	<b>0.011</b>	38 (23.5)	0.286	
Mini-GDS≥1	251 (32.8)	75 (31.6)	0.738	78 (39.0)	0.100	
MNA≤23.5	480 (57.4)	181 (69.9)	<b>&lt;0.001</b>	183 (80.6)	<b>&lt;0.0001</b>	
CIRS-G≥1 comorbidity grade 3/4	270 (56.6)	85 (60.7)	0.387	80 (66.1)	0.058	
TUG≥20 s	177 (40.4)	57 (44.5)	0.405	44 (41.5)	0.836	
Screening scores						
G8	Normal >14	173 (20.5)	28 (10.7)	<b>&lt;0.001</b>	17 (7.5)	<b>&lt;0.0001</b>
	Abnormal ≤14	671 (79.5)	233 (89.3)		211 (92.5)	
	Median (IQR)	11 (8.5;14)	11 (8.5;13)	<b>0.009</b>	9 (7.5;12)	<b>&lt;0.0001</b>
Modified G8	Normal <6	181 (21.4)	27 (10.3)	<b>&lt;0.0001</b>	17 (7.5)	<b>&lt;0.0001</b>
	Abnormal ≥6	663 (78.6)	234 (89.7)		211 (92.5)	
	Median (IQR)	17 (7;25)	19 (10;25)	<b>0.040</b>	21 (14;27)	<b>&lt;0.0001</b>

Abbreviations: IQR, interquartile range; ECOG-PS, Eastern Cooperative Oncology Group-Performance Status; ADL, activities of daily living; IADL, Instrumental ADL; MMSE, Mini Mental State Examination; Mini-GDS, Mini Geriatric Depression Scale; MNA, Mini Nutritional Assessment; CIRS-G, Cumulative Illness Rating Scale for Geriatrics; TUG, timed up-and-go test.



**Figure A8.1.** Kaplan-Meier curves of overall 3-years survival for the 6 main tumor sites



**Titre : EVALUATION DE LA FRAGILITE EN ONCOLOGIE GERIATRIQUE : DEVELOPPEMENT ET VALIDATION D'UNE NOUVELLE ECHELLE DE DEPISTAGE.**

**Mots clés :** fragilité, sujet âgée, cancer, dépistage, développement, validation, analyse de survie.

**Résumé :** Une évaluation gériatrique (EG) est recommandée pour tous les patients âgés atteints de cancer pour identifier d'éventuels problèmes de santé et ainsi optimiser la stratégie thérapeutique. Néanmoins, elle est très consommatrice de temps et de moyens. Plusieurs outils de dépistage ont été développés mais a) aucun ne dispose de propriétés diagnostiques adéquates en pratique clinique, b) ils ont le plus souvent été développés sur la base d'opinions d'experts sans développement statistique spécifique, et c) peu de données sont disponibles pour apprécier leur robustesse au changement de gold standard pour définir le concept de « fragilité ». Par conséquent, notre objectif était 1) de construire et valider un outil de dépistage performant de la fragilité et de le comparer à l'outil de dépistage G8 actuellement utilisé en oncogériatrie, 2) d'en tester la robustesse vis-à-vis de 6 définitions de référence de la fragilité et 3) d'en évaluer la valeur pronostique pour la mortalité à 1 et 3 ans. Les données étaient issues de deux cohortes prospectives de patients âgés atteints de cancer : ELCAPA (2007–2012: n=729 [développement]), et ONCODAGE (2008–2010: n=1304 [validation externe]). L'outil G8 modifié final (aire sous la courbe ROC [AUROC] : 91,6% ; Sensibilité=89% ; Spécificité=79%) comprenait 6 items indépendants : perte de poids, problèmes neuropsychologiques, statut fonctionnel, état de santé perçu, poly-prescription et existence parmi les antécédents d'une insuffisance cardiaque ou coronaropathie. Les travaux accomplis dans le cadre de cette thèse ont de plus permis de confirmer ses bonnes propriétés diagnostiques en validation externe (AUROC : 84,6% ; Sensibilité=82% ; Spécificité=69%), sa robustesse au changement de gold standard et sa valeur pronostique forte vis-à-vis de la mortalité. L'utilisation du G8 et G8 modifié devrait être encouragée en oncologie gériatrique.

**Title : ASSESSMENT OF FRAILTY IN GERIATRIC ONCOLOGY: DEVELOPMENT AND VALIDATION OF A NEW SCREENING TOOL**

**Key words:** frailty, older person, cancer, screening, development, validation, survival analysis.

**Abstract:** The comprehensive geriatric assessment (GA) is recommended in older cancer patients to inventory health problems and tailor treatment decisions accordingly. However, GA is time- and resource-consuming. Several screening tools have been developed but a) their diagnostic performance is insufficient, b) most instruments have been developed exclusively on the basis of experts' opinions without any specific statistical psychometric development, and c) a wide variability of criteria have been used to define "frailty" as the gold standard, with no investigations of their influence on the diagnostic properties of screening instruments. Therefore, our objective was 1) to develop and validate a new screening tool for frailty that achieves high diagnostic accuracy, and to compare it to the G8 screening tool, currently used in oncogeriatrics, 2) to evaluate its robustness to modifications on the gold standard, for which 6 reference definitions were tested, and 3) to assess its prognostic value for 1- and 3-years mortality. We used two prospective cohorts of older patients with cancer: ELCAPA (2007–2012: n=729 [development]), and ONCODAGE (2008–2010: n=1304 [external validation]). The final modified G8 (area under the ROC curve [AUROC]: 91.6%; Sensitivity=89%; Specificity=79%) included 6 independent items: weight loss, cognition/mood, performance status, self-rated health status, polypharmacy and history of heart failure or coronary heart disease. The work carried out in this thesis has also confirmed its good diagnostic properties in external validation analyses (AUROC: 84.6%; Sensitivity=82%; Specificity=69%), its robustness to modifications of the gold standard definition and its strong prognostic value for overall survival. The use of the G8 and modified G8 should be encouraged in geriatric oncology.