

Quality Insurance of rectal cancer – phase 3: statistical methods to benchmark centers on a set of quality indicators – Supplement part I

KCE reports 161S

The Belgian Health Care Knowledge Centre

Introduction: The Belgian Health Care Knowledge Centre (KCE) is an organization of public interest, created on the 24th of December 2002 under the supervision of the Minister of Public Health and Social Affairs. KCE is in charge of conducting studies that support the political decision making on health care and health insurance.

Executive Board

Actual Members: Pierre Gillet (President), Dirk Cuypers (Vice-president), Jo De Cock (Vice-president), Frank Van Massenhove (Vice-president), Maggie De Block, Jean-Pierre Baeyens, Ri de Ridder, Olivier De Stexhe, Johan Pauwels, Daniel Devos, Jean-Noël Godin, Xavier De Cuyper, Palstermans Paul, Xavier Brenez, Rita Thys, Marc Moens, Marco Schetgen, Patrick Verertbruggen, Michel Foulon, Myriam Hubinon, Michael Callens, Bernard Lange, Jean-Claude Praet.

Substitute Members: Rita Cuypers, Christiaan De Coster, Benoît Collin, Lambert Stamatakis, Karel Vermeyen, Katrien Kesteloot, Bart Ooghe, Frederic Lernoux, Anne Vanderstappen, Greet Musch, Geert Messiaen, Anne Remacle, Roland Lemeye, Annick Poncé, Pierre Smiets, Jan Bertels, Celien Van Moerkerke, Yolande Husden, Ludo Meyers, Olivier Thonon, François Perl.

Government commissioner: Yves Roger

Management

Chief Executive Officer: Raf Mertens

Assistant Chief Executive Officer: Jean-Pierre Closon

Information

Federaal Kenniscentrum voor de gezondheidszorg - Centre fédéral d'expertise des soins de santé –
Belgian Health Care Knowledge Centre.
Centre Administratif Botanique, Doorbuilding (10th floor)
Boulevard du Jardin Botanique 55
B-1000 Brussels
Belgium
Tel: +32 [0]2 287 33 88
Fax: +32 [0]2 287 33 85
Email : info@kce.fgov.be
Web : <http://www.kce.fgov.be>

Quality Assurance of rectal
cancer diagnosis and treatment –
phase 3: statistical methods to
benchmark centres on a set of
quality indicators – Supplement
part I

KCE reports 161S

ELS GOETGHEBEUR, RONAN VAN ROSSEM, KATRIEN BAERT, KURT VANHOUTTE,
TOM BOTERBERG, PIETER DEMETTER, MARK DE RIDDER, DAVID HARRINGTON,
MARC PEETERS, GUY STORME, JOHANNA VERHULST, VLAYEN JOAN,
VRIJENS FRANCE, STIJN VANSTEELANDT, WIM CELEN.

KCE reports 161S

Title:	Quality Insurance of rectal cancer – phase 3: statistical methods to benchmark centers on a set of quality indicators – Supplement part I
Authors:	Els Goetghebeur (UGent), Ronan Van Rossem (UGent), Katrien Baert (UGent), Kurt Vanhoutte (UGent), Tom Boterberg (UZ Gent), Pieter Demetter (Erasme), Mark De Ridder (UZ Brussel), David Harrington (Harvard), Marc Peeters (UZ Antwerp), Guy Storme (UZ Brussel), Johanna Verhulst (UZGent), Joan Vlayen (KCE), France Vrijens (KCE), Stijn Vansteelandt (Ugent), Wim Ceelen (UZgent)
Reviewers:	none
External experts:	PROCARE members: Anne Jouret-Mourin (UCL), Alex Kartheuser (UCL), Stephanie Laurent (UZGent), Gaëtan Molle (Hôpital Jolimont La Louvière), Freddy Penninckx (UZ Leuven, president steering group), Jean-Luc Van Laethem (ULB), Koen Vindevoghel (OLV Lourdes Waregem), Xavier de Béthune (ANMC), Catherine Legrand (UCL), Stefan Michiels (Institut Bordet), Ward Rommel (Vlaamse Liga tegen Kanker)
Acknowledgements:	The authors thank the PROCARE steering group, their volunteer contributors and patients for making this continued effort for quality improvement. The authors are also grateful to Alain Visscher (Ugent), Geert Silversmit (Ugent) and Carine Staessens (Ugent) for technical assistance in the making of this report. Finally the authors thank Elisabeth Van Eyck (BCR) and Koen Beirens (BCR) for their assistance on the PROCARE database.
External validators:	Pr Johan Hellings (ICURO), Pr Pierre Honoré (CHU Liège), Pr Hans C. van Houwelingen (Leiden University).
Conflict of interest:	Any other direct or indirect relationship with a producer, distributor or healthcare institution that could be interpreted as a conflict of interests: Vindevoghel Koen
Disclaimer:	<ul style="list-style-type: none">- The <u>external experts</u> were consulted about a (preliminary) version of the scientific report. Their comments were discussed during meetings. They did not co-author the scientific report and did not necessarily agree with its content.- Subsequently, a (final) version was submitted to the <u>validators</u>. The validation of the report results from a consensus or a voting process between the validators. The validators did not co-author the scientific report and did not necessarily all three agree with its content.- Finally, this report has been approved by common assent by the <u>Executive Board</u>.- Only the <u>KCE</u> is responsible for errors or omissions that could persist. The policy recommendations are also under the full responsibility of the KCE.
Layout:	Ine Verhulst
Brussels, July 12 th 2011	
Study nr 2010-04	
Domain: Good Clinical Practice (GCP)	
MeSH: Rectal neoplasms ; Quality of health care ; Quality indicators, health care ; Benchmarking ; Regression Analysis	

NLM classification: WI 610

Language: English

Format: Adobe® PDF™ (A4)

Legal depot: D/2011/10.273/41

This document is available on the website of the Belgian Health Care Knowledge Centre

KCE reports are published under a “by-nc-nd” Creative Commons Licence (<http://creativecommons.org/licenses/by-nc-nd/2.0/be/deed.en>).

How to refer to this document?

Goetghebeur E, Van Rossem R, Baert K, Vanhoutte K, Boterberg T, Demetter P, De Ridder M, Harrington D, Peeters M, Storme G, Verhulst J, Vlayen J, Vrijens F, Vansteelandt S, Ceelen W. Quality Insurance of rectal cancer – phase 3: statistical methods to benchmark centers on a set of quality indicators – Supplement part I. Good Clinical Practice (GCP). Brussels: Belgian Health Care Knowledge Centre (KCE). 2011. KCE Report 161S. D/2011/10.273/41



Appendix 1: Detailed discussion of the methodology with technical specifications and a simulation study

APPENDIX 1: DETAILED DISCUSSION OF THE METHODOLOGY WITH TECHNICAL SPECIFICATIONS AND A SIMULATION STUDY.....	1
1 INTRODUCTION.....	4
1.1 GOAL.....	4
1.2 STRATEGY.....	5
1.3 KEY FINDINGS.....	7
1.3.1 Limitations.....	7
1.3.2 Arguments for adjusting for factors such as patient Socio Economic status (SES).....	9
1.3.3 On the Instrumental Variables method	10
1.3.4 Outcome regression methods and propensity score methods	11
1.3.5 Results.....	13
2 DESCRIPTIVE STATISTICS.....	21
3 OUTCOME REGRESSION METHODS.....	23
3.1 CORRECTING FOR PATIENT-SPECIFIC COVARIATES.....	25
3.1.1 Binary outcomes	26
3.1.2 Survival outcomes: Cox proportional hazards model.....	31
4 PROPENSITY SCORE METHODS.....	35
4.1 MOTIVATION.....	35
4.2 OVERVIEW OF PROPENSITY SCORE METHODS	36
4.3 PROPENSITY SCORE METHODS FOR CENTER EFFECTS.....	37
4.3.1 Binary outcomes	37
4.3.2 Survival outcomes.....	40
5 INSTRUMENTAL VARIABLE METHODS	42
6 MISSING DATA.....	43
7 SIMULATIONS	45
7.1 PREPARING THE DATA GENERATING MODEL.....	45

7.2	GENERATING THE SIMULATED DATASETS.....	45
8	KEY POINTS	47
	BIBLIOGRAPHY	48
9	RISK-ADJUSTMENT METHODS FOR HOSPITAL PROFILING (TECHNICAL)	52
9.1	NOTATION	55
9.2	REGRESSION METHODS.....	57
9.3	PROPENSITY SCORES	68
9.4	SIMULATIONS BASED ON THE ORIGINAL PROCARE DATABASE	72
10	ESTIMATION OF CENTER EFFECTS (TECHNICAL).....	118
10.1	ESTIMATION OF CENTER EFFECTS FOR INDIVIDUAL QCI	118
10.2	ALL OR NONE QUALITY INDEX	126

1 INTRODUCTION

This first Deliverable on statistical methods for case mix adjustment for quality of care indicators is by its very nature relatively technical. Since the different methods used rely on different assumptions and can correspondingly result in a different evaluation for any given center, it is important that physician-scientists may understand the key elements involved in the modeling and are introduced to the available options. We have therefore sought to make the first part of this document accessible to a broader audience of physician-scientists. Somewhere towards Section 3 the development becomes more technical and oriented towards statisticians and epidemiologists. More detail still on these developments is provided in the technical chapter 9.

As a disclaimer at this stage we would like to emphasize that nothing shown here should be taken as an actual data analysis on the PROCARE database. In this phase we have merely simulated 'PROCARE like' data to allow us to establish performance of statistical methods in this setting.

Note: in this Deliverable we are concerned with adjusting QCIs observed in centers for the patient mix they treat, but not with benchmarking or setting standards of care. The latter will be the object of study in Deliverable 3.

1.1 GOAL

The goal of the current study is to develop the methodology to identify low and high performing hospitals in the management of rectum cancer, on the basis of the available set of QCI. **The methodology developed will be generic and applicable to other cancers.**

Our charge for this Deliverable 1 is to develop a method that allows adjusting QCI measures per center for the patient mix treated by the center so as to ultimately arrive at one or more global quality indexes with well understood benchmarks. This adjustment for the patient mix is anticipated to be most important in the outcome rather than process domain since in principle process QCIs have by definition been adapted to the patient type where needed. As part of our charge we will also examine whether a more parsimonious set of indicators could achieve a similarly effective feedback result. Fewer QCIs to register may encourage participation, reduce missing data and involuntary measurement error. Since the current charge was launched, the PROCARE steering group has revised its original set of QCIs, proposed 11 new ones, deleted 3 and adapted several existing ones as described in Appendix 3. The

new set of QCI's can be derived from data in the current PROCARE database without need to link to external databases.

1.2 STRATEGY

To reach the goal of identifying low and high performing hospitals in the management of rectum cancer (RC) on the basis of the available set of quality of care indicators (QCI), we first translated the question within a conceptual and operational framework.

The framework most relevant here is that of causal inference: we wish to evaluate not just an association between centers and outcomes, but the effect caused by hospital, over and above the patient characteristics, on the patient's treatment quality or outcome. In other words, we aim to find out what would happen if a well defined group of patients were treated by provider A rather than provider B. For this purpose we wish to first correct for patient-specific characteristics but not for hospital characteristics since those are considered part of the package the hospital brings to the patient. Once this correction exercise is completed, we will turn to hospital-specific characteristics which may help explain any variation in center effects and thus perhaps point to ways of improvement.

To derive a patient risk adjusted measure of hospital performance, the project aimed to have access to data from two cohorts; the smaller more comprehensive PROCARE database as well as an administrative (claims) database. The original 40 process and outcome quality of care indicators can be derived from the combined data in those databases and further information is available there on the patients background and general health, which may be prognostic for the treatment process and outcome QCI's. As the project got launched, however, the PROCARE steering group refused coupling of the PROCARE database with other existing databases for this goal. As a result, some of the original QCI's are no longer measurable and few baseline covariates remain. We do have access to clinical baseline variables. The former aspect is largely remedied through the proposed updated set of QCI's. The problem of substantially limited access to potential confounders appears much more serious. It has led to some modification of the methodological development plan and will ultimately weaken its application in this setting as described in the next Section.

At both levels of the analysis, special attention will go to center sizes which are known to vary substantially. At the first level, we will need to consider that centers which provide data on just a few patients produce a very weak evidence base for the center's general effect measurement. If the few patients have been selected among more, they carry the additional risk of some selection bias. Confidence/credibility

intervals on the center-specific QCI summary may then be so wide as to be non-informative and cover regions of excellence, as well as of average and poor performance. Random effects models and/ or Bayesian models are designed to overcome this in part by borrowing information from an assumed population distribution of center effects.

Center size may have a further impact beyond the precision of our estimates. For instance, high volume centers are likely specialized and hence perhaps subject to a more complicated case mix and could have better or worse comparative performance for that very reason. For the purpose of evaluating center-specific quality of care, we do not plan to adjust for center-specific covariates but see them as part of the center package just like other center-specific covariates. Hence in its potential role of prognostic factor, center size will only enter the analysis in the second round. Equally, any interaction effects between center and patient-specific covariates, would indicate that similar patients fare differently in different centers. For instance, a center specialized in geriatric medicine may care particularly well for older rectum cancer patients. We will not control for this in the primary analysis but will explore such mechanisms in the second round, when we are explaining differences seen in (patient mix adjusted) center performance.

With the above considerations in mind we consider three main methods for risk-adjustment:

1. Standard outcome regression methods (ORM), adjusting for available confounders and possibly incorporating random center effects.
2. Methods using the propensity score (PS), this is the estimated probability that a patient with a given set of risk factors was treated in each of the considered hospitals.
3. Instrumental variable (IV) methods where the IV, i.e. a predictor for the hospital which is not further predictive of the outcome, is used as a vehicle to estimate the hospital effect.

The vast majority of the measured QCIs are binary measures. In addition there are several important right-censored survival time measures (to be summarized in for instance overall 5-year survival probability, the relative survival and the disease-specific 5-year survival probability). Beyond this, there is a QCI describing the number of lymph nodes examined, which would perhaps most naturally be approached as a continuous or count measure, but can equally be treated using survival methodology (since the e.g. number of lymph nodes is positive a semi-

parametric model with multiplicative effect of covariates on the intensity of lymph nodes examined may reasonably be fitted). Since treatment of continuous outcome measures tends to be the most straightforward, methodologically speaking, we will concentrate in this text on the development for binary and survival type outcomes. We observe at this point, that the QICs for 5-year survival will not be mature in the PROCARE database that will be made available, which is restricted to patients diagnosed since 2006 and followed up until the start of 2010. In our implementation, we will therefore focus on x-year survival with x the maximum possible, given the limited data. $X = 2$ years for the preliminary database received and will likely be 3 years for the updated database we are to receive.

1.3 KEY FINDINGS

Before entering into detail on the methods, we lay out here our general findings and options taken, which are further supported by developments in the text below as well as in an extensive technical chapter (9). We thus set out to consider three classes of methods from the most standard to the most state-of-the-art for risk adjustment in the evaluation of causal effects: from outcome regression methods over propensity score methods to instrumental variables methods. We conducted our evaluation considering both the general assessment of quality of care and the specific context of the PROCARE database and the data structure (to be) made available to us.

The first two approaches (ORM and PS) rely on the assumption of ‘no unmeasured confounders’ for estimation of the (causal) effect of center on quality outcome. In contrast, the instrumental variables approach allows for unmeasured confounders but requires an instrumental variable instead: a variable which is associated with center but not otherwise with the natural outcome of the patient. Important limitations in light of these requirements result from the restricted access to baseline data in the PROCARE database which include for instance age, gender, C-staging at diagnosis and ASA score for co-morbidity (on a 4 point scale), but no access to such variables as

1. socio-economic status (SES),
2. specific co-morbidity, or
3. patient distance from the treatment center.

1.3.1 Limitations

We briefly explain the limitations entailed by missing 1.-3. and the methodological choices resulting from that. The three variables mentioned are representative of

different types of information not directly available in the PROCARE database, but potentially available through linking with other existing databases such as the IMA database.

1. SES represents a variable which is possibly a confounder for the center-quality relationship through the link with a specific natural risk profile (over and beyond what is contained in age-gender-C-staging), while it may at the same time influence a treatment quality, irrespective of the center, for instance because patients in a higher SES stratum more easily receive a more expensive or specific treatment [1]
2. Specific co-morbidities could definitely change the risk profile and would justify or may even require an adapted treatment.
3. Distance, or some derived measure thereof such as distance to a given center relative to the nearest center distance, is likely a strong predictor of center choice, and could be an instrumental variable if it does not further affect the quality outcome. In several instances in the literature a measure of distance, location or region was proposed in this sense [2-8]. Alternatively, if distance affects outcome because of its association with region and perhaps a particular local toxin or genetic form of the cancer, or if it moderates treatment - for instance through reduced visits with a longer distance, or the choice of a closer center when more frequent visits are required - it is a confounder or mediator and not an instrument.

So, first all three variables 1.- 3. could be confounders, that is, a common cause of center choice and outcome quality, for which one needs to adjust if the pure center effect is to be measured. Second, both SES and co-morbidity may generate a different treatment response for otherwise similar patients (across all centers). In an optimal quality setting SES should not influence treatment while co-morbidity should. In light of this, some scientists feel one should not adjust for SES when analyzing treatment effects in view of benchmarking. We argue that in a practical setting where SES does influence treatment across the board (for all centers) the most relevant effect measure for the patient as well as the most fair comparison of quality delivered by centers is obtained after adjusting the effect measure for SES. The arguments for this are summarized in Subsection 1.4.2.

Third, if distance between patient and treatment center influences the treatment (schedule) received and hence outcome, it affects outcome directly and can no longer serve as an instrumental variable. The general implications of all three points for our analysis approach are described following the next Subsection.

1.3.2 Arguments for adjusting for factors such as patient Socio Economic status (SES)

Background: a host of patient-specific characteristics (at diagnosis) influences the outcome of rectum cancer patients. Not all of these factors are known or can be carefully measured. Currently we are adjusting for just a few pre-treatment patient-specific factors, including age, gender, C-staging of the cancer at diagnosis, possibly ASA score, etc.. The implication is that we predict risks of individuals based on limited prognostic information and then see how the observed risk in a center deviates from that. The question is, should we or should we not in principle also adjust for such factors as SES if we can (potentially obtained through a link with the IMA database), knowing that in practice:

1. different SES may be treated differently across all centers: higher SES gets a more expensive and better treatment element [1], say, and
2. different SES patients may present themselves with different natural progression because of distinct environmental, genetic, co-morbidity conditions beyond what has been measured through C-staging, ASA-score etc. in a necessarily limited prospective voluntary register.

Without adjustment we fail to correct for a possibly associated differential natural risk (which is always needed) as well as for SES-related differences in treatment (which we may or may not wish to adjust for if conditional on SES the treatment adaptation happens irrespective of the treatment center). With adjustment, we adjust for both different risk levels and different treatment levels associated with SES and hence do not penalize centers who carry a heavier load of the 'worse treated patients'.

Conclusion: If our perspective is the one of the patient: 'given who I am, where should I go to get the better treatment/outcome' then the most relevant answer would be found after adjusting for SES. This is true whether or not we evaluate the centers for the population of their own typical patient mix or for a fixed population average outcome. Hence one should adjust for SES (like) factors if at all possible, to get the more scientific and relevant answers as well as an honest comparison of *differential* performance between centers.

If we would simply wish to alert the center to the fact that it has worse outcomes than other centers (which may be due to its different patient mix which may or may not be well treated) then an unadjusted analysis is in order. Since our primary goal in this deliverable is on adjusting for patient mix, we will adjust for SES whenever possible, even though unadjusted reports have their own contribution to make.

As we are unable to adjust the analysis for some known confounders, we must acknowledge that patient adjustments constructed (by regression and the propensity score method) will only partially correct and the residual center effects defined may result in part from differential representation of these factors in the center's patient mix. Whether or not this is the case, can only be examined once the additional set of covariates becomes available for analysis.

The propensity score approach might be weakened as the distance, a likely strong predictor of center, cannot be included in the propensity score. This would be a special point of concern when the distance is also moderately associated with the outcome, for then it is an important confounder, although not otherwise.

1.3.3 On the Instrumental Variables method

For the combined set of reasons stated below, we will not use instrumental variables in this project.

- Lacking the measures on the patients distance to every center considered we are unable to involve it in the analysis as an instrumental variable. No other potential instrumental variables were recovered based on the literature search from Deliverable 2.
- If distance is associated with outcome or treatment (schedule), either because the schedule gets adapted to the distance or the other way around, instrumental variable property is violated and it becomes an invalid instrument.
- Preliminary results indicate that the presence of that many centers with a correspondingly small propensity makes that there is too little information about the causal effect of the centers if one wishes to allow for unmeasured confounders. This is translated into confidence intervals so wide they become unusable.

Even though the instrumental variables approach is unworkable in the current setting, there may be a future role for it. While we cannot recognize the actual identity of specific centers and hence have no direct information on center type, it is clear that certain centers differ from others in important aspects. For instance, University hospitals tend to differ in size (larger), in equipment and staff they can draw on (more state of the art, costly, highly trained) and in the population they attract (more difficult cases). As a cluster they tend to draw on more resources which would suggest they have their own standard to aspire to. They are centers specifically dedicated to the advancement of science and its implementation in practice. It might be worth having a secondary analysis of center effects confined to this cluster of fewer and larger centers, for the development of their own benchmark. Here the argument of tiny propensity scores would vanish and distance could again become a workable instrument on the condition the instrument is rich enough to avoid multicollinearity in a two stage regression and no serious confounding or mediation through the distance remains.

1.3.4 Outcome regression methods and propensity score methods

For our goal, we now focus on the outcome regression methods and propensity score methods in more detail. Notwithstanding the limitations in the current setting, both approaches have their merit here and more generally when the full scale of confounders and prognostic factors for center choice are included in the analysis.

To arrive at a meaningful evaluation and the comparison of outcome regression and propensity score methods, several basic choices are made. Different methods concentrate on direct modeling of distinct target parameters. These involve patient-specific, center-specific or population-specific risk estimation. Patient-specific adjustments are the more standard direct focus of modeling and will form building blocks of our models. Here, population-specific risks express risk of a certain event if all patients in one chosen common study population were treated in a given center. In contrast, center-specific measures compare the observed risk for patients in a given center with the risk that these same patients would have experienced in some 'average' center. Evidently, from the measures conditioning on more detailed information the more averaged measures can always be derived, but not the other way around. It was found that center-specific treatment effects are best evaluated on the patient mix they themselves currently treat. Hence this will be our primary aggregated outcome measure, even though this means that different centers are judged on different patient mixes. This reference was seen to be particularly relevant

in a stable landscape where the patient mix tends not to change much over the years. Drastic interventions in the treatment landscape could of course make this stability premise untrue.

The center-specific treatment effect will most easily be derived from outcome regression models (fixed or hierarchical). Current implementation of a (fixed effect) propensity score method naturally focuses on population averaged effects only. As indicated, such an effect measure has the great advantage that it constitutes a common reference outcome for all centers and can be derived from the results of all methods. Our comparisons of results of different approaches in this report will examine both measures before coming to a conclusion in this report. While a propensity score based matched analysis can in principle be developed, this is documented to be less reliable than what we obtain through the doubly robust propensity based methods, a version of the method which protects against misspecification of either the outcome regression model or the propensity score model for center choice, and will therefore not be pursued here.

Either approach and target parameter leaves the question: relative to which 'specific center' effect do we express our adjusted outcome measures? There are (at least) two basic options studied in Deliverable 3: an external (international) reference or standard, and an internal (to the PROCARE data base) reference. Here we briefly discuss the latter only – in view of the modeling choices to be made. The discussion on benchmarking and quality standards is left to Deliverable 3. Standard regression models involving a separate effect for each center in addition to the effects of patient-specific characteristics parameterize center deviations from either

- a single chosen reference center (the first, last, largest, best, or on a percentile) - through 'dummy coding'
- the average center effect, averaged over all centers (on the given scale) - through 'unweighted effect coding' or
- the average center effect, averaged over all patients - through 'weighted effect coding'.

With weighted effect coding, large centers get more weight in defining the reference which is not the case with unweighted effect coding.

With those choices in mind we have developed a number of modeling options below. We study in detail the fixed effect outcome regression, random effects outcome regression and a doubly robust propensity score method. We focus here on models

for the most important, most common as well as most challenging outcome types which are binary outcomes (success) and right censored survival type outcomes (time to event). As prototype cases we focused on outcome QCI 1111 (overall observed survival) and QCI 1232a (proportion of APR and Hartman procedures among patients who underwent radical surgical resection). Their theoretical properties were considered and – more importantly - their practical potential performance in the PROCARE setting was evaluated through simulation based on preliminary data made available to us on August 4, 2010. The simulations are deemed necessary because the presence of small centers (some with just a single patient entered) precludes an uncritical reliance on asymptotic properties of model parameter estimators and, a fortiori, of estimators of center-specific effects. Through a well chosen computational data generating mechanism, the simulations allow one to study the accuracy of a particular method in a particular setting before implementing it there.

The precise set-up of the simulations is given in later Sections of this document and in more detail in the technical chapter. Basically, they mimic the available database and first generate a random center choice in function of base line characteristics based on a propensity score. Next, from the chosen center a random outcome is generated for the patient based on the outcome regression model. It is thereby assumed that center effects are themselves randomly distributed with some variation over the various centers in the database. Because the propensity scores are fitted on the original data, they reflect also the variation in center size seen in the database.

After fitting the various models, we display when possible both the estimated center-specific effects and population averaged center effects for the different centers in our preliminary database. Based on the repeated simulations we get insight in the variation of the estimators as they vary from simulated dataset to simulated dataset. We are concerned specifically with bias, precision and coverage of confidence intervals. We further consider center-specific risks and population averaged risks estimated over all centers.

1.3.5 Results

In this section we outline basic results for the binary QCI 1232a (proportion of APR and Hartman procedures among patients who underwent radical surgical resection). More detail and further results, including on the survival outcome, can be found in the “Technical Chapter”.

Figure 1 shows boxplots of estimated effects on the available preliminary PROCARE dataset. The first two show estimated center-specific chances of the QCI 1232a, using fixed effects logistic regression model (Firth-corrected) and a hierarchical logistic regression model assuming a normal distribution of the center effects. The final three show estimated population averaged chances of QCI 1232a, first for these same two methods and then for the propensity score method.

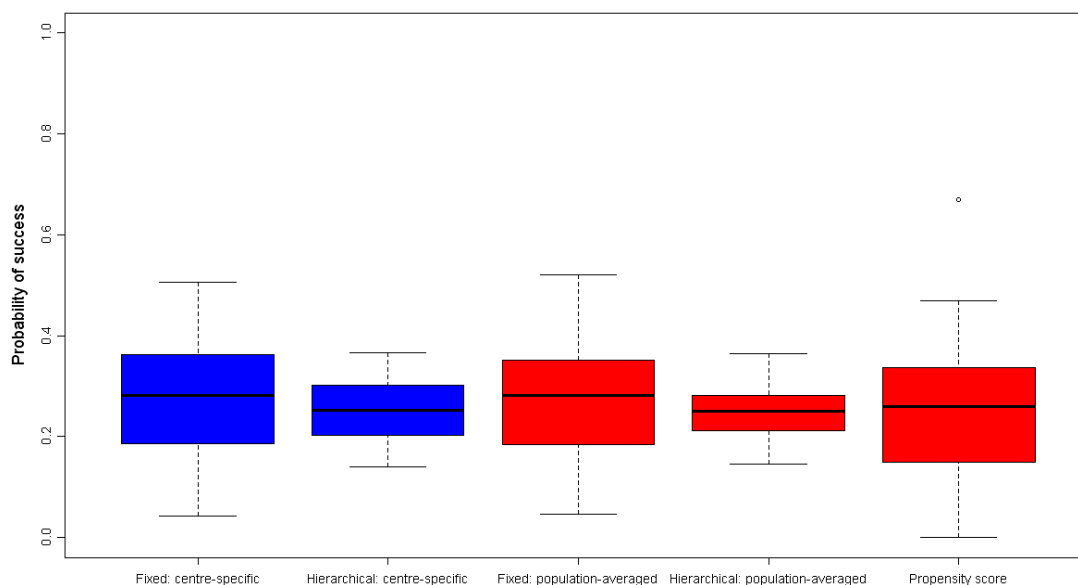


Figure 1: Comparison of estimates produced by the different statistical methods for the probability that QCI 1232a (proportion of APR and Hartman procedures among patients who underwent radical surgical resection) is present - on the original preliminary PROCARE dataset. 'Fixed' stands for the (Firth corrected) fixed effects logistic regression model, 'Hierarchical' for the hierarchical logistic regression model with normal random effects model and 'Propensity score' for the regression double robust estimator (involving a standard logistic regression).

Hierarchical models show a narrower spread when estimating the same distribution of center effects. This is an expected consequence of the fact that their estimated effect sizes are shrunk towards the center average combined with the fact that some extra information is brought in through the assumption of modeled effect distribution across centers. A key question is whether the extra spread produced by the other methods reflects just extra noise (imprecision or random error on the estimates) or genuine extra variation in the true center effects. Part of the answer is found through simulations which we have performed and show for each method how the true center

effects (represented by red triangles) have been estimated over the different simulated databases. Below we show this for the population averaged chances estimated under the hierarchical logistic regression and propensity score model. The horizontal lines on each graph show the 95 % most central chance estimates produced for each center along with the average estimate which shows up as a blue bullet. Ideally the blue bullet (average estimate) and red triangle ('true', i.e. simulated, parameter) are quite close, and the narrower the width of the interval the less variable our estimates are. Figure 2 shows clearly how the shrinkage and narrower estimation intervals for the hierarchical model sometimes completely misses the 'true', i.e. simulated, center effect. This is a well documented feature of the method. In Figure 3 the median width of the corresponding intervals for the doubly robust propensity score method is approximately three times as long but the estimates turn out to be well centered around the target parameters for each center.

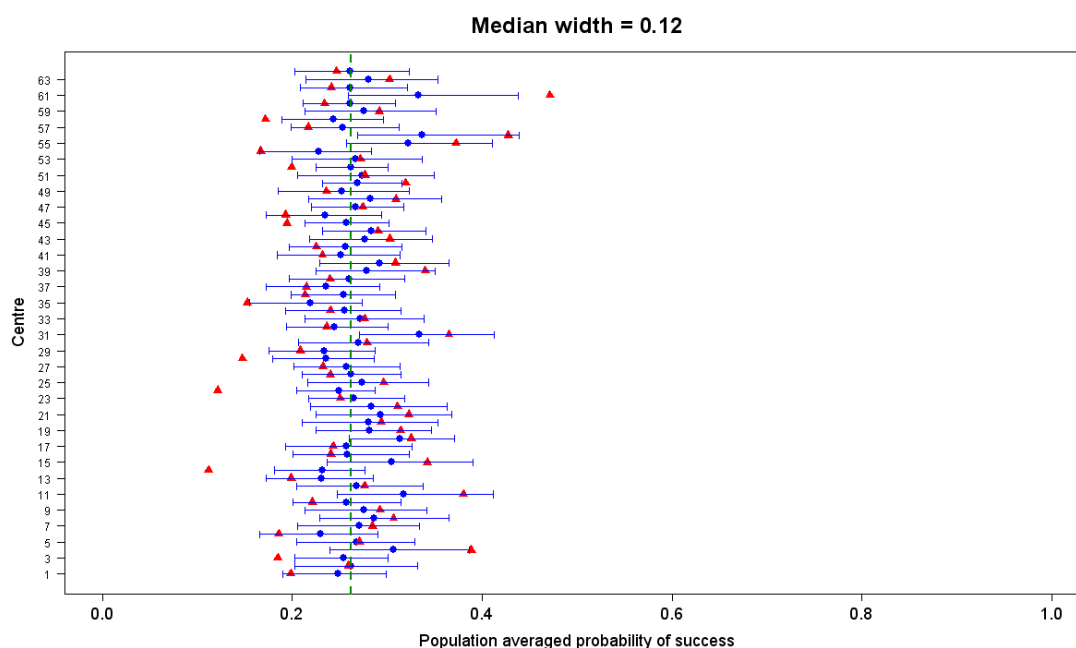


Figure 2: Estimation of the population-averaged probability of success with QCI 1232a on the simulated datasets through the *hierarchical logistic regression method*. For each center, the red triangles represent the 'true' population averaged probabilities of success, the blue bullets represent the average of the correspondingly estimated probabilities of success over the 1000 simulations and the intervals show the range of the 95% central estimates, they are thus based on the empirical distribution of all simulated population averaged probabilities of success.

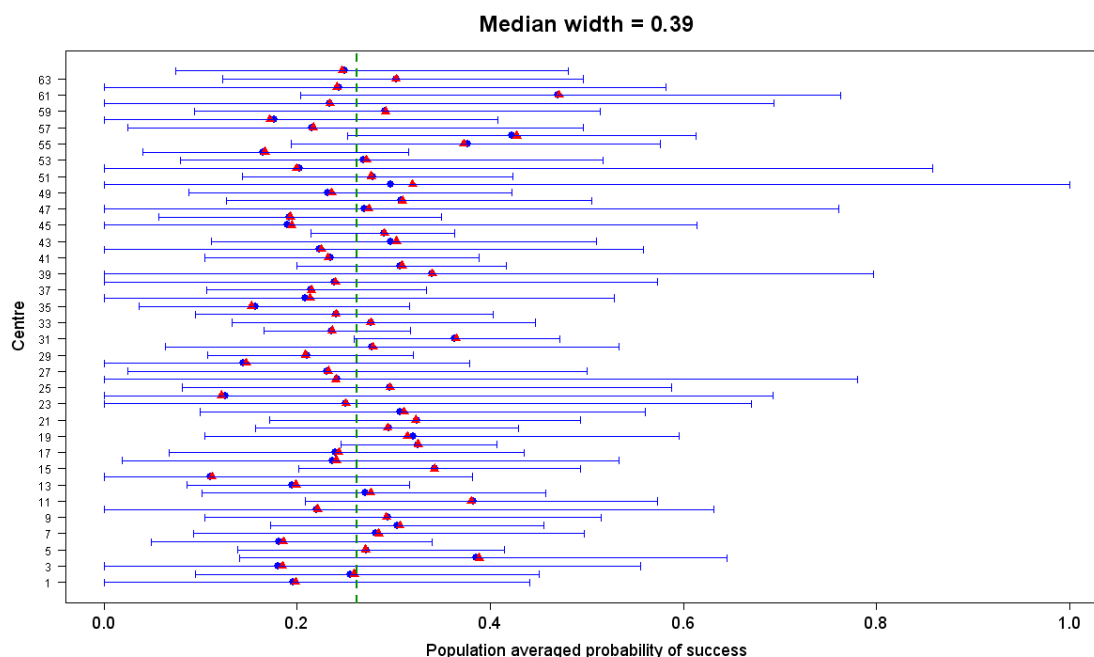


Figure 3: Estimation of the population-averaged probability of success with QCI 1232a on the simulated datasets through the *propensity score method*. For each center, the red triangles represent the ‘true’ population averaged probabilities of success, the blue bullets represent the average of the correspondingly estimated probabilities of success over the 1000 simulations and the intervals show the range of the 95% central estimates, they are thus based on the empirical distribution of all simulated population averaged probabilities of success.

To get an indication of the coverage of estimated 95% confidence intervals in this setting, we centered the depicted intervals around each of the separate estimates and verified for each center for what percentage of the simulated datasets the resulting confidence interval covered the truth. This yields a distribution of coverage estimates over the centers for each method as shown in table Table 1. This measure is complemented by the median width of the empirical 95% confidence intervals, as a measure of efficiency. A third measure of how well the estimators perform is given by the root mean squared error, which is like a standard deviation of the estimates, but centered around the truth rather than the average estimate. This is shown in Table 2. Considered together, these measures point to a choice of estimator.

Table 1: Distribution over the centers of the observed coverage of the 95% empirical confidence intervals, and median width of the intervals when estimating QCI 1232a success rates using a normal random effects model for the normally distributed random effects.

Method	Outcome measure	Min	25%-tile	Median	75%-tile	Max	Median width
O/E method	Standardised event rate (SER)	0	14	62	85	96	1.20
	Odds ratio	68	87	91	93	95	2.07
Fixed-effect	Centre-specific probability of 'success'	72	85	88	91	95	0.37
logistic regression	Population-averaged probability of 'success'	70	89	92	94	96	0.31
	(Centre-specific) standardised odds ratio	69	90	93	94	95	1.71
	Odds ratio	0	67	89	94	96	0.63
Hierarchical	Centre-specific probability of 'success'	1	76	91	94	95	0.14
logistic regression	Population-averaged probability of 'success'	0	74	91	94	96	0.12
	(Centre-specific) standardised odds ratio	0	66	89	94	96	0.61
Propensity score	Population-averaged probability of 'success'	72	92	94	95	95	0.39

Table 1 reveals how in terms of median width of the 95% empirical confidence intervals, this is the width of the central 95% range of the estimates, the hierarchical model produces the shortest intervals and the propensity score method the longest. For the commonly targeted parameter, population-averaged probability of QCI 1232a 'success', those widths are 12% and 39% respectively, it is 31% for the fixed-effect logistic regression method with Firth correction. The short intervals of the hierarchical logistic regression come at a price in terms of coverage. In the 25% centers with the lowest coverage, for instance, the true center effect was covered for the hierarchical logistic regression by no more than 74% of the 'empirical 95% confidence' intervals, while coverage was found in those centers to reach 89% for the fixed-effect logistic regression estimates with Firth correction and 92% for the propensity score method. Note how the minimum values in the first column point to some (small) centers for which the true target was never covered with the hierarchical logistic regression estimates. In summary, coverage is best achieved by the propensity score method, followed closely by the fixed-effect logistic regression method with Firth correction.

In contrast, when the focus is on root mean squared error, the hierarchical logistic regression model wins, with the propensity score method the runner up as shown in Table 2.

Table 2: Distribution over the centers of the root mean squared error of the estimated parameter describing QCI 1232a success rates using a normal random effects model for normally distributed random effects

Method	Outcome measure	Min	25%-tile	Median	75%-tile	Max
O/E method	Standardised event rate (SER)	0.195	0.451	0.623	0.848	2.675
Fixed-effect logistic regression	Odds ratio	0.219	0.384	0.546	0.717	∞
	Centre-specific probability of 'success'	0.056	0.089	0.103	0.134	0.210
	Population-averaged probability of 'success'	0.049	0.076	0.089	0.110	0.189
	(Centre-specific) standardised odds ratio	0.194	0.335	0.446	0.603	∞
Hierarchical logistic regression	Odds ratio	0.116	0.159	0.236	0.336	1.445
	Centre-specific probability of 'success'	0.027	0.037	0.044	0.059	0.152
	Population-averaged probability of 'success'	0.026	0.035	0.041	0.055	0.146
	(Centre-specific) standardised odds ratio	0.114	0.157	0.233	0.330	1.368
Propensity score	Population-averaged probability of 'success'	0.038	0.076	0.104	0.133	0.293

In general we found that – as expected – fixed-effects logistic regression estimates are more variable and hence show wider confidence intervals than their random effects counterparts. This is true, even after a Firth correction was used in the fixed effects model, penalizing the likelihood as explained in Section 3.1.1.2 to avoid exploding standard errors due to a complete separation (no residual variation) of outcomes in the small centers. The hierarchical models, even when they are implemented with the correct random effects distribution model that actually generated the data, may well produce more accurate estimators for some of the center's effects, but equally fail to detect outlying centers in more instances than we would hope to see. The technical Chapter explores its further properties under the 'less favorable' scenario where data are generated following a *bivariate* normal distribution ignored by the data analysis model, which still works with a single normal distribution. With the team, we agreed to produce both estimates (fixed- and random effects) for estimation of center-specific effects and population averaged center effects as illustrated in Figure 4. On those occasions where they disagree about the qualitative assessment of the performance of a center, a more in depth look will be necessary accounting for the differences in performance of the estimators as outlined above and in the "Technical Chapter".

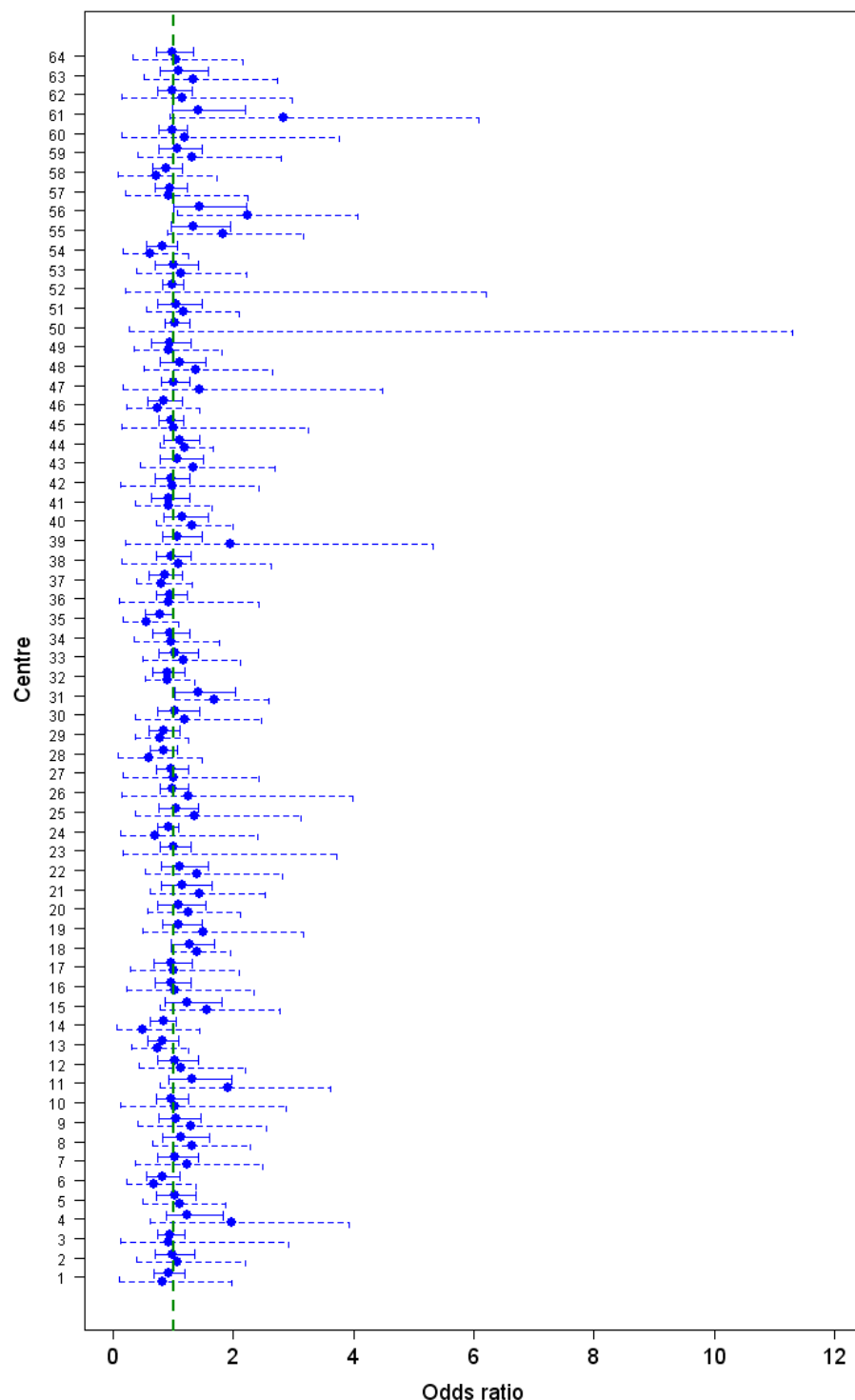


Figure 4: Forest plot to illustrate shrinkage. Blue dots represent the average of the simulated odds ratio's from the *fixed-effects* and *hierarchical logistic regression model* and empirical 95% confidence intervals for the fixed-effects logistic regression model are represented with a dotted line (--) and for the hierarchical logistic regression model with a full line (-).

To make comparisons with the propensity score method, we are currently confined to the estimated population averaged effects, which they target directly and which can be derived from the patient-specific estimates of the outcome regression models. In the setting above we found the coverage of the estimates and derived confidence intervals to be the most accurate. This comes at a price in terms of precision: we found the widest confidence intervals for the propensity score method. This method however enjoys a robustness property that makes it particularly attractive when model building becomes harder with a large number of covariates to adjust for. In such setting the method could also regain precision as further explained in Section 4. Again, given the known properties and available evidence there is no reason to claim a uniformly better or worse performance of this method.

We note two further points on the doubly robust propensity score method, which is much less tried in this setting. There is no theoretical reason why a Firth correction could not be implemented along with it. We plan to do this in the future. Theoretically too, the method could be expanded to yield center-specific estimates. While feasible in principle, such development has not been tried before and is therefore considered outside the scope of this project.

In summary, with regard to the center-specific effects which are not estimated by the standard propensity score methods, we see no reason to distrust estimated center effects with confidence intervals for the fixed effects models, but could benefit for some centers substantially from the tighter random effects estimates when the model is correct. The team decided that the PROCARE evaluation is well served by a visual display of both estimates (fixed and random effects) with corresponding confidence interval for each center. For population averaged effects, a comparison with the propensity score method results, which do not rely on the outcome regression model being correct will also be prudent and worthwhile. In many cases the same qualitative conclusions will result from the different evaluations. If and when they do differ a more in depth examination will be required in the specific setting.

Finally, results under a misspecified random effects model are shown in the "Technical Chapter". They are rather encouraging and largely follow the lines above. For right censored survival data with a focus on 2 year survival, results are more tentative due to few events in a sizeable number of centers. When 3 year survival becomes available in an updated dataset, we will be able to draw more firm conclusions for that setting. The expectation is that these will broadly follow the lines just described for the binary data.

2 DESCRIPTIVE STATISTICS

Before embarking on more complex modeling, descriptive statistics on outcomes, centers, and prognostic factors is good statistical practice and will help define the scope of analysis. Due regard is to be given to missing data at this level. While we are not planning to elaborate on the standard approach to this in any detail here, we simply point to some more important features to be examined in our setting.

For key survival outcomes, examination of the distribution of follow-up time in the dataset and over the centers, together with the observed numbers of events will give an indication of the amount of information in the dataset and each of the centers. It will for instance reveal whether 5-year survival chances are estimable with any degree of confidence, given the extent of follow-up. If updated yearly, such measures per yearly epoch may also yield a helpful description of the center progress over time in response to the monitoring and feedback. Further for this outcome type, it is important to consider whether censoring is or appears to be non-informative, possibly conditional on certain factors, before embarking on any analysis. If censoring is related to observed covariates, conditioning on those factors will be necessary in (cause-specific) survival models to avoid censoring bias. Alternatively, marginal survival models can be fitted in combination with methods for dependent censoring which involve these covariates [9]. Depending on the event (cause-specific or not) Kaplan-Meier Survival curves or the cause-specific cumulative incidence curves will non-parametrically describe the proportion of patients avoiding specific events over time.

A similar basic description of other QCI is warranted: tables for discrete (binary) variables, boxplots, and summary statistics for continuous outcomes and counts.

Regarding the centers, a first descriptive analysis should shed light on the variation in center size and the percentage of very small centers for which negligible information may be available. Secondly it will be important to recognize whether centers differ in amount of follow-up time (and therefore the censoring distribution) as well as more general completeness (missing data) over the centers. Finally, especially for sizeable centers, a brief inspection of covariates and correlation between covariates can help reveal whether some forms of center-specific characteristics, suggest special selection or measurement error and could be further examined. Detailed data quality control and a study of possibly systematic selective patient recording lies however in the hands of PROCARE and the Belgian Cancer Register who, unlike ourselves, have access to important background data in this regard (such as what percentage of

its patients the center actually registered in the PROCARE database, and how the profile of its registered patients differs from that of those patients it did not register). This is beyond the scope of the current project and we will hence proceed with methods assuming we are dealing with a relevant sample of the observed patient population over the given treatment centers.

Finally, we will examine the distribution of patient characteristics observed in the database and over the centers. Again, missing data patterns, measures of location and variation plus correlation between and among QCI as well as their prognostic factors could vary substantially between centers. This will reveal, among other things, the importance of adjusting for specific characteristics in the patient mix. If there turns out to be little or no overlap however, the adjustment for those covariates based on a general model fit may no longer be meaningful [10].

3 OUTCOME REGRESSION METHODS

Here and in the Sections to follow we give some more detail on the general methods that we consider using in this setting. Section 3 is concerned with methods that are more standard generally and in this field and will therefore be less detailed.

Our primary goal, described first in this development, is to understand how the distinct hospital centers differ in outcome they tend to produce for similar patient populations. Since in our observational data, the patients seen in different hospitals may differ in terms of their risk factor (distribution), and since we do not wish to confound the hospital effect with the effects of these patient-specific characteristics, we will adjust for them when regressing QCI on center. On the other hand, specific hospital attributes which may affect QCIs/outcomes for patients beyond what is expected based on their own characteristics, are part of the package the hospital offers and will not be taken out of the total effect equation by conditioning on these characteristics. In a second instance we will however seek to explain any differences seen at the first level in terms of hospital-specific factors such as the size of the hospital, the size carried by its surgeons, comprehensiveness of the service offered, type of treatment (schedules) they tend to work with, ... In the context of PROCARE hospital-specific confounders would not be used as an 'excuse' for a potentially lower QCI but may point to ways of improvement. The directed acyclic graph (DAG) [11] in Figure 5 summarizes the relation between the variables described above for causal effect estimation of center on QCI through regression; the dotted line, representing a causal relation between the hospital choice and the QCI is of main interest here. The set of possibly measured confounders for the hospital choice and QCI contain:

- Patient-specific confounders, a re e. g. a ge, gender, C -staging a t diagnosis, socio-economic status (SES), co morbidities, ...
- Hospital-specific prognostic factors, such as hospital volume, surgeon case l oad, p rocess or ganization, nu mber of nurses, t reatment preference, ... which are seen as part of the package that constitutes the center effect. We do not adjust for then in our primary analysis.

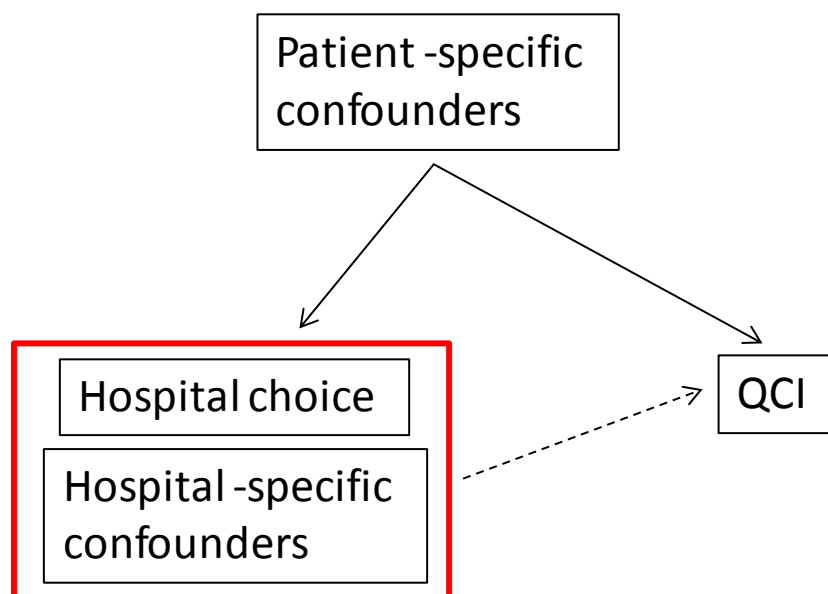


Figure 5: Directed acyclic graph (DAG) for the regression context, where hospital choice and hospital-specific confounders are considered as one.

Our analysis will start from the premise (assumption) that there are no *unmeasured* patient-specific confounders for hospital choice and QCI. In view of limited availability of prognostic factors we may need to ultimately enter into a sensitivity analysis acknowledging unmeasured confounders with levels and impact suggested by the literature search as provided in Deliverable 2. Causal theory then indicates that, in order to estimate the ‘pure’ causal effect of the hospital on the expected QCI (relative to some well defined reference), the regression analysis should correct for (i.e. be conditional on) all patient-specific confounders.

The resulting residual center effect expresses how far the expected hospital QCI deviates from what is expected under the reference conditions, based on its patient mix. Once hospital (relative) specific effects are measured in this way, a next goal is to explain the corresponding variation between hospitals in terms of observable hospital characteristics.

This result can in turn lead to constructive suggestions for improving the quality of care in all hospitals treating rectal cancer patients.

To achieve this secondary goal we will regress the QCI on both hospital-specific prognostic factors and patient-specific confounders and thus estimate the direct causal effect of interest. For both stated goals above, we will present several types of regression models and discuss feasibility, underlying assumptions, interpretation issues, ...

3.1 CORRECTING FOR PATIENT-SPECIFIC COVARIATES

We aim at estimating the causal effect of the hospital on binary (, continuous) and right-censored QCI, to then use these estimates to benchmark hospitals based on their performance for the specific QCI or a global quality index, and eventually to explain the estimated differences in performance based on the hospital-specific covariates. To this end we explicitly consider hospital choice and hospital-specific covariates as one 'package' and decide to only use hospital-specific information to explain differences in the modeled performance indices.

Each type of QCI, binary (,continuous) and right-censored survival, require an adapted modeling strategy. Binary outcomes are typically analyzed using a logistic regression model, continuous outcomes using a linear regression model and survival outcomes most often using a Cox proportional hazards model. A separate Section is dedicated to binary and survival outcomes. Most literature on provider profiling discusses and analyzes binary outcomes only. They consider that since the hospital effects of interest are estimated at the same level as the patient-specific prognostic effects for all hospitals [10], it is important to have sufficient overlap in patient populations between the hospitals for this comparison to be meaningful. This is implicit in the adjustment for case-mix.

At this stage it is worth mentioning that all fully parametric methods developed in this document allow for a Bayesian as well as frequentist definition with corresponding estimation procedure. So far, we have emphasized the frequentist approach but brought in a Bayesian-like element through the Firth correction. The Bayesian methods have the advantage that they are not concerned with asymptotic properties of estimators and allow for a very flexible transformations of estimated parameters following a MCMC implementation. They have also the well known drawbacks that 1) prior knowledge on the model parameters must be provided, 2) results and conclusions rely on the prior distributions as well on the (correct specification of the) parametric models, 3) estimation is computer intensive if MCMC is used whereby an extra element of randomness and subjective decisions enters the evaluation and conclusions and 4) frequentist properties of estimators may be unknown. The converse is of course then true for frequentist methods. Especially when they rely on asymptotic (near) normality a critical evaluation of their small sample properties will be required in the specific setting.

In what follows, we will refer to the following two definitions:

Regression-to-the-mean bias: Describes the tendency for institutions that have been identified as 'extreme' to become less extreme when monitored in the future – put simply, part of the reason for their extremity was a run of good or bad luck. This simple phenomenon could lead to spurious claims being made about the benefit of interventions to 'rescue' failing institutions. Shrinkage estimation (in hierarchical models) is intended to counter this difficulty of 'false positive' findings. [12]

Shrinkage: Individual hospital-effects are shrunken toward the mean intercept. This effect occurs in an analysis using hierarchical models, especially when the heterogeneity between the hospitals is large and the observed effect is down-weighted for high volume hospitals.

3.1.1 Binary outcomes

For now we will focus on the logistic regression approach for binary outcomes which is needed for most of the QCI's. We distinguish three methods for analyzing binary outcomes using a logistic regression model:

1. O/E method (indirect standardization through logistic regression)
2. Fixed-effect logistic regression
3. Hierarchical logistic regression

Technical details of these methods are described in Chapter 9.

In the text below we act as if the binary QCI is an indicator for mortality, but terminology can of course be adapted appropriately according to the meaning of the QCI.

3.1.1.1 *O/E method: Indirect standardization using a fixed-effect logistic regression model*

We start from the logistic regression model with only patient-specific confounders and compute the 'expected mortality rate', which may equally be an 'expected event rate' or 'expected success rate', depending on the nature of the event which is indicated by '1' rather than '0'. The ratio of the observed mortality rate in a hospital over the expected mortality rate in that same hospital is called the *standardized mortality rate* (SMR). An elementary assessment of the performance of a hospital is to compare its SMR and corresponding 95% confidence intervals with 1 [13]. Those hospitals whose 95% confidence intervals lie entirely below 1 are classified as low outliers and those hospitals whose 95% confidence intervals lie entirely above 1 are classified as high outliers. In other areas of science one is more concerned with a

given magnitude of effect before labelling an outcome as an outlier. This discussion is however deferred until deliverable 4

A related measure is the *risk adjusted mortality rate* (RAMR) which is simply computed as the product of the SMR and the overall mortality rate (over all hospitals). An elementary assessment of the performance of a hospital is to compare its RAMR and corresponding 95% confidence intervals with the overall mortality rate [14]. Those hospitals whose 95% confidence intervals lie entirely below the overall mortality rate are classified as low outliers and those hospitals whose 95% confidence intervals lie entirely above the overall mortality rate are classified as high outliers.

While this approach is simple, and provides a useful descriptive tool it has some drawbacks.

Assumptions: Conditional on the centre-specific probabilities of mortality, the observed outcome indicators are assumed to be mutually independent. This does not actually hold because the within-hospital correlation cannot be ignored.

Precision and accuracy are harder to derive when expected and observed outcome are derived from the same database. Simulations and resampling methods can shed light on this.

Interpretation of results: Easy interpretation, even for non-statisticians.

Feasibility: Very feasible

Shrinkage: No shrinkage

Ability to detect outliers: Good [15] but should be examined for different scenario's (e.g. sample sizes) through simulation.

Handling different sample sizes: All hospitals are treated similarly, there is no special correction for different sample sizes.

Multiple testing: Selection of extreme centers based on this measure implicitly involves multiple testing with its dangers of false positives.

Bayesian versus frequentist approach: A corresponding Bayesian method has been developed to estimate the Bayesian RAMR [15] and was suggested it for future use as it avoids approximations inherent in the frequentist inference method. The estimated values of RAMR and Bayesian RAMR are essentially identical and identical outliers are detected, but the intervals are quite different, they do of course also aim to cover different quantities.

3.1.1.2 *Fixed-effect logistic regression*

Rather than computing an SMR or RAMR for each hospital it is possible to estimate all hospital effects (always relative to a reference hospital) in a fixed-effect logistic regression model. [16] warn users that different coding schemes for the hospital effects can influence the ranking substantially. We are however not so concerned with ranking. They considered so called 'effect coding' and 'weighted effect coding', and point out that dummy coding is not of interest since firstly it produces exactly the same results as effect coding and secondly it does not allow comparing one hospital to an overall mean contrary to effect coding. The choice of a single reference center with dummy coding would also appear quite arbitrary.

Assumptions: Independence of binary outcomes conditional on the centre-specific probabilities of mortality. While we cannot adjust for post treatment variables, given our goal, such variables may explain extra correlation within centers and GEE like methods would allow to account for this at the variance level.

Precision and accuracy: Relatively few assumptions are made and correspondingly wide confidence intervals. No shrinkage of effect estimates for small centers

Interpretation of results: Interpretation in terms of odds ratio's relative to the chosen 'reference hospital' and at the same value for the patient-specific covariates in the model. Using effect coding avoids the arbitrary choice of a specific hospital as the reference and allows to achieve more precise estimates. Also, with effect coding the reference result is shifted toward rates of small providers when the quality of care measure is related to hospital volume. This could explain some large inconsistencies seen between the O/E method and fixed-effect models [16]. Based on this model, estimation of the mortality rate at a given hospital, is most directly derived at a specified value of the patient-specific covariates. For center-specific or population based averages, some further averaging is needed.

Feasibility: Might not converge if there are centers with only a few patients (which is to be expected). In fact, this was found to be a problem on our preliminary database. In response, we found how Firth's correction for small centers allows to reduce this problem and further reduces bias in the process [17]. While the correction failed to converge in reasonable time for our data set in R (version 2.10.1), it did work well in SAS (version 9.2). Simulations for this method were therefore moved to SAS. Additionally, we have chosen to limit individual center analysis to centers with at least 5 registered patients. The smaller centers will be grouped together and carry the special label.

Shrinkage: There is no shrinkage for the standard logistic regression, but some shrinkage when the Firth correction is added.

Ability to detect outliers: Depends partly on the used coding scheme for the centers. [close to that in the previous approach]

Handling different sample sizes: There is no implicit correction in the standard application, but some with Firth's correction. The different sizes do influence the reference value from which deviations are measured through weighted versus unweighted effect coding.

Bayesian versus frequentist approach: The use of Firth's correction in the frequentist analysis reduces the gap between the frequentist and Bayesian approach in this setting. Indeed, the correction consists of a penalty added to the score equation to be solved. The penalty term involves Jeffrey's invariant prior used in a standard Bayesian analysis [17].

3.1.1.3 *Hierarchical logistic regression*

Rather than modeling the hospital-effects explicitly in a fixed effects model several authors suggest implementing a hierarchical (also called random intercept- or multilevel) logistic regression model with two levels:

- First level (within-hospital or patient-level): model the probability of the QCI in function of patient-specific characteristics.
- Second level (between-hospital or hospital-level): model the variation of the log OR across the hospitals, one speaks of random effects (or of frailties in the survival setting).

For two-level structured data, although the hierarchical model allows dependence among patients within hospitals, it does assume the independent random sampling of hospitals and hence exchangeability: the joint distribution of the treatment effects is independent of the identity of the actual centers being considered. In practice, the exchangeability assumption involves two components. First, that the odds ratios are unlikely to be similar. Second, that there is no *a priori* reason to expect the odds ratio in any specified center to be larger than the odds ratio in another specified trial. This has the consequence that an *a priori* ranking of the effect sizes is not possible. [18]

Advantages over fixed-effect logistic regression:

- Structured to accommodate dependency within hospitals [19] – it has this in common with the fixed effects model, but...

- requires smaller within-hospital sample sizes, provided there is an adequate number of providers [20].
- The hierarchical model mimics the hypothesis that underlying quality leads to systematic differences among true hospital outcomes [10].

Assumptions: Exchangeability of hospitals (unless hospital-specific parameters are modeled at the second level of the model). An implicit assumption in the hierarchical logistic regression model is that hospital outcome is independent of the number of patients treated at the hospital [10]. Finally, there is of course the form of the assumed model for the between-center effects. If needed this form can be allowed to be quite complex and flexible [21].

Precision and accuracy: Due to the shrinkage phenomenon hospital-specific performances (e.g. odds ratios) are closer to the mean (one) compared to the previous two methods.

Interpretation of results: Interpretation in terms of odds ratio's relative to the chosen reference hospital or relative to the average of the other hospitals. Hierarchical modeling is efficient in the sense that the profiling estimator can be obtained directly from the model.

Feasibility: Computationally intensive if estimated in a Bayesian manner.

Shrinkage: One feature of hierarchical modeling is that estimates of the level-2 random term tend to shrink towards the mean 0. Shrinkage will be negligible when the overall within-hospital variation is negligible, but when the variation in mortality within hospitals becomes more substantial, shrinkage will be stronger. Regression-to-the-mean is naturally accommodated because posterior estimates of the random intercepts, or functions of the random intercepts are "shrunk" toward the mean [20] and [22].

Ability to detect outliers: Due to the shrinkage of 'extreme' hospitals this hierarchical model is more conservative for detecting outliers than the fixed logistic regression model.

Handling different sample sizes: Implicit shrinkage of outcome measures for small centers towards the grand mean.

Multiple testing: Multiplicity of parameter estimation is addressed by integrating all the parameters into a single model, for example, a common distribution for the random intercepts [10]. It does avoid the convergence problems of the standard logistic regression.

Bayesian versus frequentist approach: From [15] it appears that the frequentist method classified an outlying hospital that was not classified as such by the Bayesian approach. The conditions under which these discrepancies occurred have been examined and it appears that when the frequentist estimate is near 0, then the frequentist and Bayesian estimate are essentially the same and the frequentist intervals are a little larger. For the largest frequentist estimates, the (symmetric) frequentist intervals are narrower and contained within the corresponding Bayesian intervals. They prefer the Bayesian method since it does not require symmetrical intervals. Modern day frequentist methods are however no longer confined to normal asymptotic inference. Likelihood ratio tests are preferred over Wald tests in this setting. Furthermore, with resampling based methods more exact inference becomes possible.

[21] present a flexible random effects model based on methodology developed in the Bayesian non-parametrics literature. Their approach is applied to the problem of hospitals comparisons using routine performance data, and among other benefits provides a diagnostic to detect clusters of providers with unusual results, thus avoiding problems caused by masking in traditional parametric approaches. They provide code for Winbugs in the hope that the model can be used by applied statisticians.

3.1.2 Survival outcomes: Cox proportional hazards model

The statistical literature on provider profiling based on survival outcomes is limited. Since several important QCI are survival outcomes: overall 5-year survival by stage (KCE 2008 QCI 1111), relative survival (new QCI), disease-specific 5-year survival by stage (KCE 2008 QCI 1112) and disease-free survival (new QCI), we develop this in some detail here.

Data are available for patients with rectum cancer (RC) incidence dates between January 2006 (start active input into the database) and currently 31/12/2008, to be updated to include 2009. Mortality data are collected from the mortality database of the sickness funds (IMA), no mortality data are available for patients with private insurance (PROCARE II: maximal 9 out of 1071). Therefore, the survival is probably slightly overestimated. Besides this, the majority of censoring occurs due to administrative reasons (end of study – or rather closure of the mortality database) or also because patients are treated abroad or because they do not have a social security number or postal code.

We briefly discuss three methods for adjusted for covariates when analyzing survival outcomes, which gradually involve more assumptions:

- Kaplan-Meier estimation stratified by hospital and C-staging
- Cox proportional hazards model
- Cox frailty model

All of these methods rely on the assumption of non-informative censoring. This may require conditioning on center, say, if center turns out to be a predictor for the censoring distribution as well as the outcome. If the survival model does not condition on such covariates, special techniques need to be invoked to handle explainable informative censoring [23]

Technical details of these methods are described in a “Technical Chapter”.

3.1.2.1 Kaplan-Meier estimation stratified by hospital and C-staging

Since adjustment for case-mix in the different hospitals to be profiled is essential and we expect few events per stratum in each hospital, estimating the stratified Kaplan-Meier curves for each separate center involves more imprecision than can reasonably be useful here. We will use this tool as a global descriptive measure (across all centers).

3.1.2.2 Cox proportional hazards model

In Cox’s proportional hazards model we allow for a baseline (cause-specific) mortality rate with nonparametric evolution over time, and model the proportional effect of patient-specific characteristics and center on top of this (i.e. mortality or disease-specific mortality). This happens by multiplying the baseline hazard with the exponential of a linear function of the predictors [9]. Patients who survived (the event of interest) during the observation period are censored on the last day of the observation period.

In terms of advantages and disadvantages as well as pros and cons this follows the lines of the fixed effects logistic regression model. A Firth correction, to avoid non-convergence with complete separation due to small center sizes, is available here too in SAS (version 9.2) [24]. The key distinction with logistic regression is that the amount of information and hence precision is now a function of the number of observed events (and hence person years of observation) per center, rather than just the numbers of patients registered per center. Adjustments for important covariates

(like C-staging) can now also happen through stratification and hence need not be constrained by strict assumptions (such as proportional hazards over the C-stages) .

Assumptions: After adjusting for base line co variates, the hazards of the different centers are assumed to be proportional (over time) to one another (unless one stratifies on a co variate or allows for time-dependent co variates). As in the binary case, while we cannot adjust for post treatment variables, given our goal, such variables may explain extra correlation within centers and GEE like methods could allow to account for this at the variance level.

Precision and accuracy: Depends on the number of observed events per center, and hence also on the observed total person years.

Interpretation of results: Interpretation can be cast in terms of hazard ratio's relative to the chosen reference hospital or relative to the average of the other hospitals, or x-year survival can be derived from the hazard functions, for a given level of patient-specific characteristics. Effect coding may be used to have the center average hazard as the baseline hazard.

Feasibility: May not be feasible if hospitals are low-volume to the extent that no variation in outcome is (likely) observed. With small center sizes the model may not fit and standard errors become infinite. We have chosen to limit individual center analysis to centers with at least 5 registered patients. The smaller centers will be grouped together and carry the special label. The equivalent of the Firth correction can be used to overcome this in this setting [24].

Ability to detect outliers: This may depend on which point of the survival curve we are targeting. Otherwise similar to fixed effect logistic regression.

Handling different sample sizes: In no differential fashion unless the Firth correction is used [25]

3.1.2.3 *Cox frailty model*

To allow small centers to draw some information from the general distribution of outcomes, the distribution of center effects could be modeled on the hazard ratio scale. We speak of a frailty term coming from the frailty distribution instead of the fixed binary variables indicating each specific center. This is the equivalent of the random effects logistic regression model, but now for right censored time to event outcomes.

Assumptions: After adjusting for base line co variates, the hazards of the different centers are assumed to be proportional (over time) to one another. A frailty

distribution is specified for the random factor. The majority of studies assume gamma or lognormal distribution [26]. Because of the latency of the frailty term and possible sparseness of events it is generally difficult to determine an appropriate frailty distribution for a specific data set. The literature on this topic is also rather sparse [27]. As frailty models are conditional models, the proportional hazards assumption only holds conditionally on the frailties.

Precision and accuracy: Through the assumption of a shared parametric frailty distribution shrinkage occurs of extreme event rates in small centers. This is appropriate if the center is exchangeable with other centers. With the shrinkage further come narrower confidence intervals which are reliable in large samples if the specified frailty distribution turns out to be correct. Some caution is needed since a correct frailty distribution cannot be guaranteed and the power to detect deviations from an assumed one tends to be rather limited. With shrinkage estimators, the variance of the random effects will consistently underestimate the variance [28-29]. It has also been documented that because of the shrinkage, it becomes harder to detect small centers which are outlying.

Interpretation of results: Frailties quantify the heterogeneity in time to event rates between centers [26]. As for the fixed effects model, interpretation can be cast in terms of hazard ratio's relative to 'the average' of the other hospitals as implied by the mean 1 standardization of the frailty. Here too, x-year survival can be derived from the hazard functions, for a given level of patient-specific characteristics. Centers with a high frailty value perform poorly. The frailty model has the advantage that it provides a measure of the spread of outcomes over centers.

Feasibility: May not be feasible if very small low-volume hospitals are used.

Shrinkage: The frailty terms are shrinkage estimators as they are constrained by a penalty function added to the log-likelihood which tends to shrink them towards the mean [29-31].

Ability to detect outliers: Plotting the realized frailties coefficients can reveal outliers. These can be found as well by checking the martingale residuals [31-32].

4 PROPENSITY SCORE METHODS

4.1 MOTIVATION

In the previous Section, we have described statistical methods to adjust for differential case mix which are based on regression models for the association between each quality indicator on the one hand, and patient characteristics on the other hand, within each center. These methods are very powerful, but have a number of limitations in view of which propensity score methods have been developed. Since these methods have been less tried in this setting, in this Section, we will first provide insight into the motivation for considering such methods, as well as into their own potential limitations.

An important limitation of outcome regression methods primarily arises when patient characteristics are very different between centers. This is because these methods essentially attempt to compare patients with the same characteristics between different centers. When different centers have a very different case mix, then the amount of information available for making such comparisons is limited. In that case, problems of multicollinearity arise whereby the correlation between different predictors in the model (e.g. between center and patient characteristics) is so large that their own separate effects are difficult to disentangle and thus unstable estimates with large variance are obtained. It is common practice to alleviate such multicollinearity problems by simplifying the regression model, e.g. by deleting certain patient characteristics from the model. This happens essentially automatically upon applying model selection strategies (e.g. forward, backward or stepwise regression) because the large imprecision affecting regression coefficients of predictors that are subject to multicollinearity, is often a primary decision basis for deleting such predictors.

With a focus on a 'causal' center effect, such model simplification strategies can be sub-optimal for various reasons. First, when different centers have a very different case mix, then due to lack of information, the statistical analysis becomes heavily sensitive to correct specification of the model, for which goodness-of-fit tests have very limited power under these circumstances. Second, the default strategy of retaining center – because it is our primary focus – in the regression model, may lead one to systematically delete patient characteristics that are strongly associated with center choice, and thus to ascribe a possible patient mix effect incorrectly to a center effect. Third, by deleting predictors which induce multicollinearity in the analysis, one will tend to obtain center effect estimates with narrow confidence intervals. While this

may appear beneficial, a concern is that the resulting intervals leave implicit the fact that little information is available about the real center effects. In particular, it becomes very likely to obtain narrow intervals which promise to cover the population center effects with 95% chance, but in truth do not.

Most of these concerns apply primarily to settings where the number of potential confounders is large and therefore model building forms a major component of the analysis. Since the number of confounders that will be available to us in the analysis of the PROCARE data is very limited, it may turn out that model building can largely be avoided in the analysis and therefore that the aforementioned become less relevant. In this Deliverable, we nevertheless provide a thorough overview and examination of these methods for our specific setting.

4.2 OVERVIEW OF PROPENSITY SCORE METHODS

In view of the aforementioned concerns, propensity score methods [33-34] have been developed and have been found to be successful. Here, as previously explained, the propensity score refers to the probability of attending a given center in function of patient characteristics. The central idea behind most propensity score methods, which is the key result of [33], is a dimension-reduction property that all relevant patient characteristics that confound the association between center and quality indicator can be summarized into a single propensity score. This then enables the use of adjustment strategies that avoid regression modeling - and thereby overcome the previously mentioned concerns - such as matching [35] and subclassification or stratification. Also other confounding adjustment strategies like regression adjustment and inverse probability weighting based on the propensity score have been considered and will be reviewed below.

The literature on propensity scores almost exclusively focuses on dichotomous exposures and is henceforth to a large extent inapplicable to our setting where the exposure, center, is discrete with many levels. [34] proposed to focus on each paired treatment (or center) comparison, but this is not ideal for our purposes where the interest does not naturally lie in paired comparisons. Others [36-37] subclassify or regress the outcome of interest on a so-called multiple propensity score (also referred to as a propensity function in [38]). This is the vector of probabilities of attending each center, given patient characteristics, as may be obtained based on the fitted values from a multinomial regression model.

Unfortunately, also this approach is not workable for our purposes because the multiple propensity score is high-dimensional - in fact, given the many centers, it is of

even higher dimension than the set of available patient characteristics. This makes that subclassification approaches will suffer from sparse strata, that matching strategies will have difficulties finding subjects who are alike in terms of the multiple propensity score, and that regression adjustment for the multiple propensity score will suffer from over-fitting. In the following Sections, we will propose more feasible strategies, first for dichotomous outcomes and later for survival outcomes.

4.3 PROPENSITY SCORE METHODS FOR CENTER EFFECTS

4.3.1 Binary outcomes

When the quality indicator is a dichotomous event Y , e.g. mortality (coded to be 0 or 1), then our focus is on the population-averaged risk, i.e. the mortality risk that would have been observed had all patients in the study population been treated at a given center c . [39] develops inference for this probability based on the so-called generalized propensity score. Here, for a given patient, this is the probability of that patient attending his/her observed health-care provider in function of the available patient characteristics. In particular, [39-40] demonstrate that the population-averaged risk for given center c can be estimated using the following 2-step approach:

- Regress outcome Y on the generalized propensity score amongst patients in center c , e.g. by means of a logistic regression model;
- Average the fitted values from this outcome prediction model over all subjects in the sample, but with the generalized propensity score substituted with the probability of each subject attending center c , given his/her subject characteristics.

When the sample size per center is small, then one may instead use the following related approach:

- Regress outcome Y on the generalized propensity score amongst patients and center using the data from all centers;
- Average the fitted values from this outcome prediction model over all subjects in the sample, but with center set at c and with the generalized propensity score substituted with the probability of each subject attending center c , given his/her subject characteristics.

A major advantage of these approaches based on the generalized propensity score over those described in the previous Section is that the generalized propensity score

is univariate. Working with a univariate propensity score avoids the difficulties that we previously alluded to, of working with a high-dimensional multiple propensity score. The generalized propensity score brings the added merit that it will reveal to what extent some centers cannot directly be compared to certain other centers due to non-overlapping patient populations. Indeed, with many confounders available, it can be difficult to evaluate whether different centers have similar patient populations in terms of all these confounders. Since all confounders can be reduced into a univariate generalized propensity score, it suffices to evaluate whether different centers have overlap in terms of this propensity score.

A limitation of the foregoing approaches is that they rely on correct specification of both a propensity score model as well as an outcome regression model. In the following, we will suggest a closely related approach which poses lesser concerns for bias due to model misspecification. Just like the previous approach, it requires reliance on working models, but only assumes that one or the other is correctly specified. The first working model is a regression model for the outcome in center c (or all centers simultaneously) in function of patient characteristics; e.g. a logistic regression model. The second model is a working model for the multiple propensity score: the probability of a patient being treated in each center c in function of patient characteristics; e.g. a multinomial regression model. An estimate of the population-averaged risk for given center c can then be estimated using the following 2-step approach:

- Fit the outcome working model via a weighted regression of outcome on covariates amongst patients attending center c , with weights being the reciprocal of the generalized propensity score [41].
- Average the fitted values from this outcome prediction model over all subjects in the sample.

When the sample size per center is small, then one may instead use the following related approach:

- Fit the outcome working model via a weighted regression of outcome on covariates amongst all patients, with weights being the reciprocal of the generalized propensity score;
- Average the fitted values from this outcome prediction model over all subjects in the sample, but with center set at c .

It can be shown using similar arguments as in [41-42] that this estimator is doubly robust in the sense that it is a unbiased estimator of the population-averaged risk (in sufficiently large samples) if either the outcome regression model or the propensity score model is correctly specified, but not necessarily both.

The usefulness of doubly robust estimators has recently been questioned [43] with the argument that the performance of such estimators may deteriorate rather substantially when both working models are only mildly misspecified. In a later discussion on the paper, [41] argue that this criticism is somewhat misguided for the following reasons. First, the simulation design from which the evidence in [43] was drawn, appears to have been carefully chosen to make the doubly robust estimator perform badly. Second, the doubly robust estimators considered by [44], unlike other doubly robust estimators, are grossly inefficient when at least one of the working models is misspecified. The Kang and Schafer paper has stimulated much research on improving the performance of doubly robust estimators. The estimator that we propose here incorporates some of the latest state-of-the-art modifications designed to improve the performance of these estimators, especially in the presence of working model misspecification. In particular, unlike other doubly robust estimators, it guarantees an estimate of the population-averaged risk between 0 and 100%. Further, unlike other doubly robust estimators, it does not inflate the bias due to model misspecification in regions where the weights are large [45].

A disadvantage of using the proposed doubly robust estimator is that, when the outcome working model is correctly specified, it will be less efficient (i.e. have larger variance) than a pure regression-based estimator such as some of the estimators considered in Section 3. A further drawback is that estimates can be unstable when the weights are large for some individuals. Following a recommendation by [46], we have therefore truncated all weights at the 1% and 99% percentile in all analyses. There are also several advantages to the use of doubly robust estimators. First, they have a weaker reliance on correct model misspecification than a regression-based approach and than a pure propensity score-based approach. This may be of interest, considering the sensitivities that may be involved in benchmarking health-care centers. In particular, if centers turn out to be very different in terms of patient mix, then the doubly robust estimator will not be subject to model extrapolations unlike outcome regression-based approaches which may extrapolate the association between outcome and patient characteristics from one center to another under such circumstances. The reason that such extrapolations can be avoided is because the doubly robust estimator allows for misspecification of the outcome regression model,

in which case it relies on correct specification of the generalized propensity score. The latter merely quantifies the percentage of patients attending one's own center at each covariate level. In such circumstances, the doubly robust estimator may have inflated imprecision, but this may merely be providing a more honest reflection of the uncertainty in the estimate which is present when different centers have a very different case mix. Second, note that inference under random effects models can be somewhat sensitive to the assumed distribution of the random effects. By using doubly robust estimators, one may share the virtues of random effect models through the outcome working model, yet have some protection against misspecification of the random effect distribution under correct specification of the propensity score model. This is likely to be promising, but has to the best of our knowledge not been studied in the literature. Finally, shrinkage bias affecting empirical estimates in random effect models may in principle compromise the validity of confidence intervals, which acknowledge imprecision, but not bias. Provided correct specification of the propensity score model, this is not the case for the doubly robust estimator, even when it involves empirical BLUPs in the outcome working model.

4.3.2 Survival outcomes

With a survival outcome, as before, our focus will be on the survival probability $S(t) = P(Y > t)$ at a given fixed point in time, e.g. 5-year survival. If there were no censoring, then estimation of the survival probability $S(t)$ would follow the lines described in the previous section. In the presence of censoring, we will rely on inverse probability of censoring weighting [47] to make progress.

Given the lack of information about the actual survival time of patients whose survival time is censored, assumptions must be made as to whether the failure rate in patients who are censored at a given time is comparable with the failure rate in patients who are not. Throughout we will allow for patients whose survival time is censored at a given point in time, to have different patient characteristics (and therefore a different survival prognosis) than uncensored patients at that time, but we will assume that all these patient characteristics are contained in the vector of patient-specific covariates on which we condition. Remember that we previously considered this set sufficient to adjust for differential patient mix. In particular, we will assume that missingness of the survival status at time t has no residual dependence on the survival status itself, given these patient characteristics. This assumption is implied by the more common assumption of non-informative censoring, following which the (cause-specific) hazard of censoring at each time t has no residual

dependence on the actual survival time (beyond time t), given the patient-specific covariates. We do not allow for the possibility that there are additional (possibly time-varying) predictors of censoring (that are also associated with survival) over and above those already contained in the considered set of patient-specific covariates. We have chosen not to do so because, in the available data, we have no access to such additional potential predictors. However, the formalism that we develop below relatively easily extends to enable these relaxations.

The inverse probability of censoring weighted estimators that we develop, rely on a working model for the probability that the survival status at time t is observed. When the focus is on a fixed time point t , then this model can be fitted using standard logistic regression. Alternatively, one may infer this probability from a hazard regression model. In addition to this model, we will – as with binary outcomes – postulate a working model for the outcome in center c (or in all centers) in function of patient characteristics.

We now propose to estimate the population-averaged probability of surviving time t in center c using the following two-step approach:

- Fit the outcome working model by a weighted regression of the survival status at time t on the patient-specific covariates within patients for whom the survival status at time t is observed and within center c (or in all centers simultaneously), with weights being the reciprocal of the product of the generalized propensity score and the probability that the survival status at time t is observed, as obtained from the censoring model. This is most easily done by using logistic regression rather than hazard regression for the outcome working model. In the simulation study the suggested weights were truncated at the 1% and 99% percentile [46] for better performance.
- Average the fitted values from this regression model over all subjects in the sample.

It can be shown that the resulting estimator is doubly robust in the sense that it is an unbiased estimator of the population-averaged survival probability at time t in center c (in large samples) if the censoring model is correctly specified and in addition, either the outcome regression model or the propensity score model are correctly specified.

5 INSTRUMENTAL VARIABLE METHODS

The methods above assume all patient characteristics simultaneously associated with the center-choice and outcome have been measured. When this is in doubt, a pseudo randomization approach can allow for unmeasured confounders provided an instrumental variable has been identified [48]. This approach requires identification of a measurable variable which predicts center, but does not predict outcome beyond that fact.

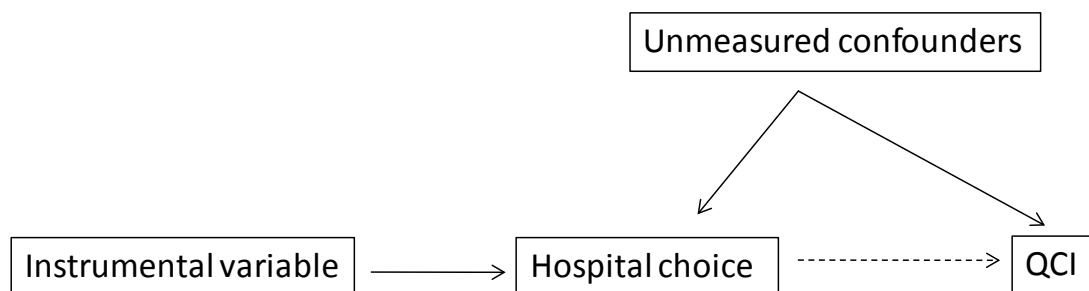


Figure 6: Directed acyclic graph (DAG) for the instrumental variable context.

Possible instrumental variables:

- For each patient the distance from home to each of the centers (multidimensional)
- The distance between the center treated at and the closest center, or rather the difference between the distance from to the center treated at and the distance from home to the closest center (one-dimensional).

Because of lack of data, possible invalidation of the IV assumption and weak information with the high number of small centers, we have abandoned this approach for the current avenue.

6 MISSING DATA

Missing data can substantially inflate the uncertainty of the study results. First, missing data mean that information that was intended to be collected, in fact was not; this reduces the sample of data that is available for analysis. Since most software routines restrict the analysis to patients for whom all data is available on the variables that are included in the analysis, this implies that even partially observed data for some of these patients can go lost. By applying state-of-the-art missing data technology, one can avoid this problem and guarantee that all available data are included in the analysis.

Second, the occurrence of missing data generates pertinent questions as to whether the subset of data on which the analysis is based, are representative of the population from which data were randomly drawn. By applying state-of-the-art missing data technology, one can allow for the missingness to be selective (e.g. for patients with missing data not to be comparable to patients with fully observed data), so long as the missingness is explainable by measured factors. For instance, if data are more likely missing for older men with early stage cancer, then the analysis can adjust for this provided gender, age and cancer staging are available. When missingness is not explainable by measured factors, but has a residual dependence on unmeasured factors, then no statistical analysis can adjust for this. In that case, sensitivity analyses must be used to evaluate how the analysis results change with varying dependence of missingness on unmeasured factors.

In the PROCARE data, missingness occurs in some of the patient characteristics (e.g. age and C-staging), as well as in some of the outcomes. Because missing ages can be approximately reconstructed from other data on these patients, the missing age problem can essentially be ignored. For sizeable missing C-staging a separate category will be used. Missing data in all remaining variables will be handled by means of sequential multiple imputation methods, also known as multiple imputation via chained equations [49-52]. Here, in the spirit of Gibbs sampling, missing data for each variable are repeatedly drawn from the conditional distribution of that variable, given all remaining (imputed) variables. The analysis is then performed on the imputed data set, which is obtained upon convergence of the algorithm. This is repeated several times to obtain multiple imputed data sets. Clever combining rules are used to combine the analysis results from these different data sets and to correct standard errors for the uncertainty regarding the imputed data.

An advantage of sequential multiple imputation relative to more standard imputation methods is that by drawing each variable separately from its conditional distribution, it can deal well with a mix of discrete and continuous measurements. A disadvantage is that there is no formal theory, which justifies the validity of this method, although simulation studies have revealed a very adequate performance.

The details of all these analyses will be made more precise as the analysis of the PROCARE data initiates, as they depend upon discussions with subject-matter experts, which will take place during the analysis phase of the PROCARE data.

7 SIMULATIONS

Below we describe in more detail the approach that was taken to arrive at simulated datasets that mimic the structure of the PROCARE database. We have done this for the binary outcome proportion of APR and Hartman procedures among patients who underwent radical surgical resection (QCI 1232a) and observed overall survival (QCI 1111). We haven't give some results for the former in Section 1.4.5 by way of illustration. For further details we refer the reader to the "Technical Chapter 9".

7.1 PREPARING THE DATA GENERATING MODEL

A hierarchical logistic regression model/frailty proportional hazards model is fitted to the available original PROCARE data, after grouping small centers (with less than 5 registered over the available period) in one overlapping 'small' center. The estimated coefficients for the fixed patient-specific characteristics (age, gender and C-staging) are stored.

The estimated variance of the random center effects/frailties is then used to generate randomly for each center one new random effect/frailty from the assumed distribution. This center effect is stored.

A multinomial propensity score model for center choice is fitted next in function of the patient-specific characteristics. From this we store for each patient the estimated chance of attending each of the centers.

7.2 GENERATING THE SIMULATED DATASETS

One thousand new datasets are randomly generated according to the database inspired model above, as follows:

- We start from the observed data on baseline covariates in the dataset. Hence the joint distribution of age-gender-C-staging is kept fixed and identical to what is in the data.
- For each patient a new center choice is randomly generated from the originally estimated propensity scores in each run of the simulations.
- Based on this new center choice and the original patient-specific characteristics, a new outcome is generated for each patient from the original model fitted with the estimated coefficients for the patient-specific characteristic, and the (once and for all datasets) generated random effects/frailties.

These one thousand datasets are then analyzed using the different techniques under study in R (version 2.10.1) and SAS (version 9.2) for the Firth-corrected analyses. Results of these analyses are stored for each generated dataset. Additionally the 'true' outcome measures for the original dataset, using the generated random effects/frailties instead of the estimated BLUPS, are obtained.

To evaluate the different methods, summary statistics are computed per center over these thousand estimated outcome measures. Additionally coverage is computed per center by applying the empirical 95 % confidence interval to each of the simulated outcomes and checking in what percentages of them the 'true' outcome measure is captured. This yields a distribution of coverages over all centers of which the minimum, median, maximum and interquartile range are computed.

8 KEY POINTS

- A more technical description of different techniques for risk-adjustment of binary and right-censored QCI is presented, considering fixed effects outcome regression, random effects outcome regression, doubly robust propensity score methods and instrumental variable methods. These four techniques are all considered within the causal framework in which we aim at estimating the effect of choice of center of care on the outcome (QCI).
- It was decided not to pursue the instrumental variables approach since the identified instrumental variables for this setting (distance and region/location) will not be available in the PROCARE database and preliminary results showed that the presence of many centers result in very imprecise estimated effects.
- An extensive simulation exercise has shown that there is no single technique that performs uniformly better than the other ones. We therefore suggest to perform all three analyses, and evaluate the combined results in light of their described strengths and limitations.
- Convergence problems when fitting simple models with center choice as fixed predictor have been identified. These problems were most prominent when small centers (with e.g. less than 5 patients) with few events were entered in the model. To ensure that the obtained results are reliable, we will restrict estimation of center effects to centers with at least 5 patients (other centers may be grouped into one overlapping center).
- Issues related to the lack of access to known confounders (e.g. socio-economic status) are discussed. The risk-adjustment analysis will necessarily be restricted to age and gender plus the baseline clinical patient-specific confounders available in the PROCARE database.
- Missing data problems have been discussed and we suggest multiple imputation techniques for reconstruction of the database under the missing at random assumption, while acknowledging that this assumption may well be violated.

BIBLIOGRAPHY

1. Olsson, L.I., F. Granstrom, and L. Pahlman, *Sphincter preservation in rectal cancer is associated with patients' socioeconomic status*. British Journal of Surgery, 2010. **97**(10): p. 1572-1581.
2. Basu, A., et al., *Use of instrumental variables in the presence of heterogeneity and self-selection: An application to treatments of breast cancer patients*. Health Economics, 2007. **16**: p. 1133-1157.
3. Brooks, J.M. and E.A. Chrischilles, *Heterogeneity and the interpretation of treatment effect estimates from risk adjustment and instrumental variable methods*. Medical Care, 2007. **45**(10): p. S123-S130.
4. Earle, C.C., et al., *Effectiveness of chemotherapy for advanced lung cancer in the elderly: Instrumental variable and propensity analysis*. Journal of Clinical Oncology, 2001. **19**(4): p. 1064-1070.
5. Hadley, J., et al., *An exploratory instrumental variable analysis of the outcomes of localized breast cancer treatments in a medicare population*. Health Economics, 2003. **12**(3): p. 171-186.
6. Sussman, J.B. and R.A. Hayward, *An IV for the RCT: using instrumental variables to adjust for treatment contamination in randomised controlled trials*. British Medical Journal, 2010. **340**.
7. Wisnivesky, J.P., et al., *Effectiveness of Radiation Therapy for Elderly Patients with Unresected Stage I and II Non-Small Cell Lung Cancer*. American Journal of Respiratory and Critical Care Medicine, 2010. **181**(3): p. 264-269.
8. Zeliadt, S.B., et al., *Survival benefit associated with adjuvant androgen deprivation therapy combined with radiotherapy for high- and low-risk patients with nonmetastatic prostate cancer*. International Journal of Radiation Oncology Biology Physics, 2006. **66**(2): p. 395-402.
9. Othuis, M., Y. Li, and R.C. Tiwari, *A Class of Semiparametric Mixture Cure Survival Models With Dependent Censoring*. Journal of the American Statistical Association, 2009. **104**(487): p. 1241-1250.
10. Normand, S.L.T. and D.M. Shahian, *Statistical and clinical aspects of hospital outcomes profiling*. Statistical Science, 2007. **22**: p. 206-226.
11. VanderWeele, T.J., M.A. Hernan, and J.M. Robins, *Causal directed acyclic graphs and the direction of unmeasured confounding bias*. Epidemiology, 2008. **19**(5): p. 720-728.
12. Spiegelhalter, D.J., K.R. Abrams, and J.P. Myles, *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Statistics in Practice. 2004, West-Sussex: John Wiley & Sons Ltd. 391.
13. DeLong, E.R., et al., *Comparing risk-adjustment methods for provider profiling*. Statistics in Medicine, 1997. **16**(23): p. 2645-2664.
14. Austin, P.C., D.A. Alter, and J.V. Tu, *The use of fixed- and random-effects models for classifying hospitals as mortality outliers: A Monte Carlo assessment*. Medical Decision Making, 2003. **23**(6): p. 526-539.
15. Racz, M.J. and J. Sedransk, *Bayesian and Frequentist Methods for Provider Profiling Using Risk-Adjusted Assessments of Medical Outcomes*. Journal of the American Statistical Association, 2010. **105**(489): p. 48-58.

16. Fedeli, U., et al., *The choice between different statistical approaches to risk-adjustment influenced the identification of outliers*. Journal of Clinical Epidemiology, 2007. **60**(8): p. 858-862.
17. Heinze, G. and M. Schemper, *A solution to the problem of separation in logistic regression*. Statistics in Medicine, 2002. **21**(16): p. 2409-2419.
18. Higgins, J.P.T. and D.J. Spiegelhalter, *Being sceptical about meta-analyses: a Bayesian perspective on magnesium trials in myocardial infarction*. International Journal of Epidemiology, 2002. **31**(1): p. 96-104.
19. Tan, A., J.L. Freeman, and D.H. Freeman, *Evaluating health care performance: Strengths and limitations of multilevel analysis*. Biometrical Journal, 2007. **49**: p. 707-718.
20. Christiansen, C.L. and C.N. Morris, *Improving the statistical approach to health care provider profiling*. Annals of Internal Medicine, 1997. **127**(8): p. 764-768.
21. Ohlssen, D.I., L.D. Sharples, and D.J. Spiegelhalter, *Flexible random-effects models using Bayesian semi-parametric models: Applications to institutional comparisons*. Statistics in Medicine, 2007. **26**(9): p. 2088-2112.
22. Normand, S.L.T., M.E. Glickman, and C.A. Gatsonis, *Statistical methods for profiling providers of medical care: Issues and applications*. Journal of the American Statistical Association, 1997. **92**(439): p. 803-814.
23. Li, Y., R.C. Tiwari, and S. Guha, *Mixture cure survival models with dependent censoring*. Journal of the Royal Statistical Society Series B-Statistical Methodology, 2007. **69**: p. 285-306.
24. Heinze, G. and L. Schemper, *A solution to the problem of monotone likelihood in Cox regression*. Biometrics, 2001. **57**(1): p. 114-119.
25. Heinze, G. and D. Dunkler, *Avoiding infinite estimates of time-dependent effects in small-sample survival studies*. Statistics in Medicine, 2008. **27**(30): p. 6455-6469.
26. Legrand, C., et al., *Heterogeneity in disease free survival between centers: lessons learned from an EORTC breast cancer trial*. Clin Trials, 2006. **3**(1): p. 10-8.
27. Balakrishnan, N. and Y. Peng, *Generalized gamma frailty model*. Stat Med, 2006. **25**(16): p. 2797-816.
28. Morris, J. S., *The BLUPs are not "best" when it comes to bootstrapping*. Statistics & Probability Letters, 2002. **56**(4): p. 425-430.
29. Duchateau, L. and P. Janssen, *Understanding Heterogeneity in Generalized Mixed and Frailty Models*. The American Statistician, 2005. **59**: p. 143-146.
30. Therneau, T.M. and P.M. Grambsch, *Penalized Cox models and Frailty*. 1998. p. 58.
31. Therneau, T.M., P.M. Grambsch, and V.S. Pankratz, *Penalized Survival Models and Frailty*. Journal of Computational and Graphical Statistics, 2003. **12**(1): p. 156-175.
32. Grambsch, P.M., T.M. Therneau, and T.R. Fleming, *Diagnostic plots to reveal functional form for covariates in multiplicative intensity models*. Biometrics, 1995. **51**(4): p. 1469-1482.

33. Rosenbaum, P. R. and D. B. Rubin, *THE CENTRAL ROLE OF THE PROPENSITY SCORE IN OBSERVATIONAL STUDIES FOR CAUSAL EFFECTS*. Biometrika, 1983. **70**(1): p. 41-55.
34. Rubin, D.B., *Estimating causal effects from large data sets using propensity scores*. Annals of Internal Medicine, 1997. **127**(8): p. 757-763.
35. Huang, I.C., et al., *Application of a propensity score approach for risk adjustment in profiling multiple physician groups on asthma care*. Health Services Research, 2005. **40**(1): p. 253-278.
36. Zanutto, E., B. Lu, and R. Hornik, *Using propensity score subclassification for multiple treatment doses to evaluate a national antidrug media campaign*. Journal of Educational and Behavioral Statistics, 2005. **30**(1): p. 59-73.
37. Spreeuwenberg, M.D., et al., *The Multiple Propensity Score as Control for Bias in the Comparison of More Than Two Treatment Arms - An Introduction From a Case Study in Mental Health*. Medical Care, 2010. **48**: p. 166-174.
38. Imai, K. and D.A. van Dyk, *Causal inference with general treatment regimes: Generalizing the propensity score*. Journal of the American Statistical Association, 2004. **99**(467): p. 854-866.
39. Imbens, G.W., *The role of the propensity score in estimating dose-response functions*. Biometrika, 2000. **87**(3): p. 706-710.
40. Imbens, G.W. and K. H. Hirano, *The propensity score with continuous treatments, in Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*. 2004, Wiley.
41. Robins, J., et al., *Performance of double-robust estimators when inverse probability weights are highly variable*. Statistical Science, 2008. **22**: p. 544-559.
42. Tsiatis, A. A., et al., *Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: A principled yet flexible approach*. Statistics in Medicine, 2008. **27**(23): p. 4658-4677.
43. Kang, J. and J.L. Schafer, *Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data*. Statistical Science, 2008. **22**: p. 523-539.
44. Kang, J.D.Y. and J.L. Schafer, *Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data*. Statistical Science, 2007. **22**(4): p. 523-539.
45. Vansteelandt, S., M. Bekaert, and G. Claeskens, *On model selection and model misspecification in causal inference*. Technical Report, 2010.
46. Cole, S.R. and M.A. Hernan, *Constructing inverse probability weights for marginal structural models*. American Journal of Epidemiology, 2008. **168**(6): p. 656-664.
47. Scharfstein, D.O., A. Rotnitzky, and J.M. Robins, *Adjusting for nonignorable drop-out using semiparametric nonresponse models*. Journal of the American Statistical Association, 1999. **94**(448): p. 1096-1120.
48. Hernan, M.A. and J.M. Robins, *Instruments for causal inference - An epidemiologist's dream?* Epidemiology, 2006. **17**(4): p. 360-372.
49. Little, R. J.A. and D. B. Rubin, *Statistical Analysis with Missing Data*. 2nd Edition ed. Wiley Series in Statistics and Probability. 2002: John Wiley & Sons, inc.

50. Burton, A. and D.G. Altman, *Missing covariate data within cancer prognostic studies: a review of current reporting and proposed guidelines*. British Journal of Cancer, 2004. **91**(1): p. 4-8.
51. Janssen, K.J.M., et al., *Missing covariate data in medical research: To impute is better than to ignore*. Journal of Clinical Epidemiology, 2010. **63**(7): p. 721-727.
52. Horton, N.J. and S.R. Lipsitz, *Multiple imputation in practice: Comparison of software packages for regression models with missing variables*. American Statistician, 2001. **55**(3): p. 244-254.

9 RISK-ADJUSTMENT METHODS FOR HOSPITAL PROFILING (TECHNICAL)

9.1 NOTATION

9.2 REGRESSION METHODS

9.3 PROPENSITY SCORES

9.4 SIMULATIONS BASED ON THE ORIGINAL PROCARE DATABASE

10 ESTIMATION OF CENTER EFFECTS (TECHNICAL)

10.1 ESTIMATION OF CENTER EFFECTS FOR INDIVIDUAL QCI

10.2 ALL OR NONE QUALITY INDEX

Risk-adjustment methods for hospital profiling

-

TECHNICAL APPENDIX

September 8, 2010

Contents

1	Notation	3
1.1	Outcome measures	3
1.2	Predictors	3
1.3	Coefficients	3
1.4	Indices	3
1.5	Definition of different outcomes	4
2	Regression methods	5
2.1	Binary outcomes	5
2.1.1	O/E method: indirect standardization using a fixed-effect logistic regression model	5
2.1.2	Fixed-effect logistic regression model	7
2.1.3	Hierarchical logistic regression model	10
2.2	Survival outcomes	12
2.2.1	Cox' proportional hazards model	12
2.2.2	Cox' frailty proportional hazards model	14
3	Propensity scores	16
3.1	Binary outcomes	16

3.2	Survival outcomes	18
4	Simulations based on the original Procare database	20
4.1	Simulation protocol	21
4.1.1	Version 1: Use correct random effects distribution	25
4.1.2	Version 2: Incorporate misspecification of the random effects distribution	25
4.2	Simulation results for the binary outcome: Proportion of APR and Hartmanns procedures (QCI 1232a)	26
4.2.1	Version 1: Use correct random effects distribution	29
4.2.2	Version 2: Incorporate misspecification of the random effects distribution	32
4.3	Simulation results for the survival outcome: Overall 5-year survival by stage (QCI 1111)	35
4.3.1	Version 1: Use correct random effects distribution	39
A	Evaluation of simulations for QCI 1232a, version 1	44
B	Evaluation of simulations for QCI 1232a, version 2	51
C	Evaluation of simulations for QCI 1111, version 1	58

1 Notation

1.1 Outcome measures

Y : continuous or binary (0/1) outcome, where a binary outcome will be coded '1' for a positive result (called 'success')

D : survival time

C : censoring time

T : observation time, i.e. $T = \min(D, C)$

Δ : status/censoring indicator

1.2 Predictors

X : matrix of dummy's for the hospitals

L : matrix of patient-specific confounders for the outcome - hospital choice relation

S : matrix of stratification variables (might be a subset of L)

Random variables are presented with capitals while observed variables are presented in lower case.

1.3 Coefficients

α : intercept

ψ : coefficients for X

β : coefficients for L

1.4 Indices

$c = 1, \dots, m$: centre indicator

$i = 1, \dots, n$: individual patient indicator or $i = 1, \dots, n_c$: individual patient indicator within centre c

$s = 1, \dots, S$: stratum indicator

$l = 1, \dots, L$: indicator for patient-specific predictors

Patient identification comes in the first index and centre identification in the second. Possible stratification factors in the third index.

Vectors are presented in bold.

Let

$$\mathbf{X}'_i = (X_{i2}, \dots, X_{im})$$

be the $1 \times (m-1)$ -vector of dummy variables for the m centres for patient i , so $X_{ic} = 1$ if patient i attends centre c

$$X = \begin{pmatrix} X_{12} & \dots & X_{1m} \\ X_{22} & \dots & X_{2m} \\ \vdots & \vdots & \vdots \\ X_{n2} & \dots & X_{nm} \end{pmatrix}$$

be the $n \times (m-1)$ -dimensional matrix with dummy variables for the m centres for all patients, and

$$\boldsymbol{\psi} = (\psi_2, \psi_3, \dots, \psi_m)'$$

be the $(m-1) \times 1$ -vector of coefficients for X .

Let

$$\mathbf{L}'_i = (L_{i1}, \dots, L_{iL})$$

be the $1 \times L$ -vector of patient-specific confounders for centre choice and outcome variables for patient i ,

$$L = \begin{pmatrix} L_{11} & \dots & L_{1L} \\ L_{21} & \dots & L_{2L} \\ \vdots & \vdots & \vdots \\ L_{n1} & \dots & L_{nL} \end{pmatrix}$$

be the $n \times L$ -dimensional matrix of patient-specific confounders for all patients, and

$$\boldsymbol{\beta} = (\beta_1, \dots, \beta_L)'$$

be the $L \times 1$ -vector of coefficients for L .

1.5 Definition of different outcomes

For continuous QCI's, Y_i (Y_{ic}) represents the observed value for patient i (in centre c).

For binary QCI's, define $Y_{ic} = 1$ if the i th patient in the c th centre received/had the QCI and $Y_{ic} = 0$ otherwise, and define

$$E(Y_{ic}|\mathbf{L}_{ic}) = P(Y_{ic} = 1|\mathbf{L}_{ic}).$$

Success (i.e. a positive outcome) will always be assigned as 1 and failure as 0.

For QCI's representing survival time, for patients i for whom the event of interest was observed, define D_i as the survival time and $\Delta_i = 1$. For patients i for whom the event of interest was not observed, define C_i as the censoring time and $\Delta_i = 0$. From this define observation time

$$T_i = \begin{cases} D_i & \text{if } \Delta_i = 1 \\ C_i & \text{if } \Delta_i = 0 \end{cases}$$

2 Regression methods

2.1 Binary outcomes

Define the logit- and expit functions as follows:

$$\begin{aligned}\text{logit}(p) &= \log\left(\frac{p}{1-p}\right) \\ \text{expit}(x) &= \frac{\exp(x)}{1 + \exp(x)}\end{aligned}$$

2.1.1 O/E method: indirect standardization using a fixed-effect logistic regression model

Start from the logistic regression model with only patient-specific confounders:

$$\begin{aligned}\text{logit}(E(Y_{ic}|\mathbf{L}_{ic})) &= \alpha + \mathbf{L}'_{ic}\boldsymbol{\beta} \\ &= \alpha + \beta_1 L_{ic1} + \dots + \beta_L L_{icL}.\end{aligned}$$

Define observed number of 'successes' in hospital c

$$O_c = \sum_{i=1}^{n_c} y_{ic},$$

and the expected number of 'successes' in hospital c

$$E_c = \sum_{i=1}^{n_c} \hat{p}_{ic}$$

with

$$\hat{p}_{ic} = \text{expit}(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\boldsymbol{\beta}}),$$

where the estimates $\hat{\alpha}$ and $\hat{\boldsymbol{\beta}}$ are obtained via standard maximum likelihood estimation.

Then the 'standardized mortality rate'(SMR) or 'standardised event rate' (SER) for centre c is computed as the ratio of the observed number of 'successes' over the expected number of 'successes' (DeLong et al. (1997)):

$$\text{SMR}_c = \text{SER}_c = \frac{O_c}{E_c}.$$

If E_c is (virtually) error-free, the 95% confidence interval for SMR_c is calculated as:

$$\left[\frac{1}{E_c} \max\left(0, O_c - 1.96 \sqrt{\sum_{i=1}^{n_c} \hat{p}_{ic}(1 - \hat{p}_{ic})}\right), \frac{1}{E_c} \min\left(n_c, O_c + 1.96 \sqrt{\sum_{i=1}^{n_c} \hat{p}_{ic}(1 - \hat{p}_{ic})}\right) \right].$$

Fedeli et al. (2007) explain, the resulting CI has been criticized because it neglects the chance variability of E , and resulted in larger CI's than are obtained from the propagation of error estimates. The latter method is better suited to the common situation where the same data set is used to develop a prediction model and to estimate expected numbers, in particular when one provider treats a large percentage of patients (Faris et al. (2003)).

An improvement of the variance calculation is provided in (DeLong et al. (1997), Appendix I), yielding the 95% confidence interval for SMR_c as follows:

$$\left[\exp \left(\ln O_c - \ln E_c - 1.96 \sqrt{\text{Var}(\ln O_c - \ln E_c)} \right), \exp \left(\ln O_c - \ln E_c + 1.96 \sqrt{\text{Var}(\ln O_c - \ln E_c)} \right) \right],$$

with

$$\text{Var}(\ln O_c - \ln E_c) = \text{Var}(\ln O_c) + \text{Var}(\ln E_c) - 2\text{Cov}(\ln O_c, \ln E_c).$$

Using the Delta method, this is approximately

$$\text{Var}(\ln O_c - \ln E_c) \approx \frac{1}{O_c^2} \text{Var}(O_c) + \frac{1}{E_c^2} \text{Var}(E_c) - \frac{2}{O_c E_c} \text{Cov}(O_c, E_c),$$

with estimated variance components:

$$\widehat{\text{Var}}(O_c) = \sum_{i=1}^{n_c} \hat{p}_{ic}(1 - \hat{p}_{ic}), \quad (1)$$

$$\widehat{\text{Var}}(E_c) \cong \sum_{i=1}^{n_c} (\hat{p}_{ic}(1 - \hat{p}_{ic}))^2 \mathbf{L}_{ic} \widehat{\Sigma}_{\alpha, \beta} \mathbf{L}_{ic}', \quad (2)$$

and

$$\widehat{\text{Cov}}(O_c, E_c) = \sum_{k=1}^{n_c} \hat{p}_{kc}(1 - \hat{p}_{kc}) \sum_{j=1}^{n_c} [\hat{p}_{jc}(1 - \hat{p}_{jc}) \mathbf{L}_{jc} \widehat{\Sigma}_{\alpha, \beta}] \mathbf{L}_{kc}' \quad (3)$$

In (2) and (3), $\widehat{\Sigma}_{\alpha, \beta}$ represents the estimated covariance matrix of the coefficients $\hat{\alpha}$ and $\hat{\beta}$ which is easily obtained from standard software or otherwise estimated as

$$\widehat{\Sigma}_{\alpha, \beta} = \mathbf{L}_{ic}' \text{diag}(\hat{p}_{i1}(1 - \hat{p}_{i1}), \dots, \hat{p}_{im}(1 - \hat{p}_{im})) \mathbf{L}_{ic}.$$

A robust version - taking into account clustering within centres - of $\widehat{\Sigma}_{\alpha, \beta}$ can also be obtained from the Design-package in R (Harrell (2001)).

Discussion points

- Instead of estimating E_c from the data, it is possible to use an external/international standard (DeLong et al. (1997)).
- Alternatively the **risk-adjusted mortality rate** (RAMR) for centre c can be computed as the SMR for that centre multiplied with the overall mortality rate (Austin et al. (2003))

$$\text{OMR} = \frac{\sum_{c=1}^m \sum_{i=1}^{n_c} y_{ic}}{n},$$

or

$$\text{RAMR}_c = \text{SMR}_c \text{OMR}$$

The 95% confidence interval for RAMR_c is calculated as

$$\left[\frac{\text{OMR}}{E_c} \max \left(0, O_c - 1.96 \sqrt{\sum_{i=1}^{n_c} \hat{p}_{ic}(1 - \hat{p}_{ic})} \right), \frac{\text{OMR}}{E_c} \min \left(n_c, O_c + 1.96 \sqrt{\sum_{i=1}^{n_c} \hat{p}_{ic}(1 - \hat{p}_{ic})} \right) \right].$$

2.1.2 Fixed-effect logistic regression model

The fixed-effect logistic regression model to be estimated looks like

$$\text{logit}(E(Y_{ic}|\mathbf{X}_i, \mathbf{L}_i)) = \alpha + \mathbf{L}_i'\boldsymbol{\beta} + \sum_{c=2}^m \psi_c X_{ic}. \quad (4)$$

For the centre effect we create $(m-1)$ variables to indicate a centre effect under two different coding schemes (instead of dummy coding):

1. Effect coding for patient i treated in centre $c = 1, \dots, m$

c	X_{i2}	X_{i3}	X_{i4}	...	X_{im}
1	-1	-1	-1	...	-1
2	1	0	0	...	0
3	0	1	0	...	0
4	0	0	1	...	0
\vdots					
m	0	0	0	...	1

2. Weighted effect coding for patient i treated in centre $c = 1, \dots, m$

c	X_{i2}	X_{i3}	X_{i4}	...	X_{im}
1	$-n_2/n_1$	$-n_3/n_1$	$-n_4/n_1$...	$-n_m/n_1$
2	1	0	0	...	0
3	0	1	0	...	0
4	0	0	1	...	0
\vdots					
m	0	0	0	...	1

Because we are using (*weighted*) *effect coding* (instead of dummy coding) for the centres it is possible to compute the odds of 'success' in centre c relative to the average of all centre-specific risks (for standard effect coding):

$$\text{OR}_{c,\text{overall}} = \begin{cases} \exp\left(-(\hat{\psi}_2 + \hat{\psi}_3 + \dots + \hat{\psi}_m)\right) & c = 1 \\ \exp\left(\hat{\psi}_c\right) & c \neq 1 \end{cases},$$

or relative to the average of all patient-specific risks (for weighted effect coding):

$$\text{OR}_{c,\text{overall}} = \begin{cases} \exp\left(-\left(\frac{n_2}{n_1}\hat{\psi}_2 + \frac{n_3}{n_1}\hat{\psi}_3 + \dots + \frac{n_m}{n_1}\hat{\psi}_m\right)\right) & c = 1 \\ \exp\left(\hat{\psi}_c\right) & c \neq 1 \end{cases}.$$

95% Wald confidence limits for these odds ratio's can be computed as follows for $c \neq 1$:

$$\left[\exp\left(\hat{\psi}_c - 1.96\widehat{\text{SE}}(\hat{\psi}_c)\right), \exp\left(\hat{\psi}_c + 1.96\widehat{\text{SE}}(\hat{\psi}_c)\right) \right],$$

or in matrix notation

$$\left[\exp\left(\hat{\boldsymbol{\psi}} - 1.96\sqrt{\text{diag}(\mathbf{X}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\psi}}\mathbf{X}')}\right), \exp\left(\hat{\boldsymbol{\psi}} + 1.96\sqrt{\text{diag}(\mathbf{X}\widehat{\boldsymbol{\Sigma}}_{\boldsymbol{\psi}}\mathbf{X}')}\right) \right],$$

with

- $\hat{\psi}$ the $m \times 1$ vector with
 - for standard effect coding: $-(\hat{\psi}_2 + \hat{\psi}_3 + \dots + \hat{\psi}_m)$ in the first element,
 - for weighted effect coding: $-\left(\frac{n_2}{n_1}\hat{\psi}_2 + \frac{n_3}{n_1}\hat{\psi}_3 + \dots + \frac{n_m}{n_1}\hat{\psi}_m\right)$ in the first element,
 and the estimated coefficients $\hat{\psi}_c$ ($c = 2, \dots, m$) in the other elements,
- $\hat{\Sigma}_{\psi}$ the $(m-1) \times (m-1)$ estimated covariance matrix of $\hat{\psi}$, and
- X the $m \times (m-1)$ matrix with a row for each centre c with corresponding coding of $X_{i2}, X_{i3}, \dots, X_{im}$,

Extra outcome measures:

- The Procure-population averaged probabilities of 'success'¹ can be estimated for each centre c as follows, for $c \neq 1$:

$$\hat{p}_c^{\text{pop.averaged}} = \frac{1}{n} \sum_{i=1}^n \text{expit}(\hat{\alpha} + \mathbf{L}'_i \hat{\beta} + \hat{\psi}_c),$$

and for $c = 1$, depending on the coding scheme:

$$\hat{p}_1^{\text{pop.averaged}} = \frac{1}{n} \sum_{i=1}^n \text{expit}(\hat{\alpha} + \mathbf{L}'_i \hat{\beta} - (\hat{\psi}_2 + \hat{\psi}_3 + \dots + \hat{\psi}_m)),$$

or

$$\hat{p}_1^{\text{pop.averaged}} = \frac{1}{n} \sum_{i=1}^n \text{expit} \left(\hat{\alpha} + \mathbf{L}'_i \hat{\beta} - \left(\frac{n_2}{n_1} \hat{\psi}_2 + \frac{n_3}{n_1} \hat{\psi}_3 + \dots + \frac{n_m}{n_1} \hat{\psi}_m \right) \right).$$

- The centre-specific probabilities of 'success' can be estimated for each centre c as follows (for $c \neq 1$):

$$\hat{p}_c^{\text{centre-specific}} = \frac{1}{n_c} \sum_{i=1}^{n_c} \text{expit}(\hat{\alpha} + \mathbf{L}'_i \hat{\beta} + \hat{\psi}_c),$$

and for $c = 1$, depending on the coding scheme:

$$\hat{p}_1^{\text{centre-specific}} = \frac{1}{n_c} \sum_{i=1}^{n_c} \text{expit}(\hat{\alpha} + \mathbf{L}'_i \hat{\beta} - (\hat{\psi}_2 + \hat{\psi}_3 + \dots + \hat{\psi}_m)),$$

or

$$\hat{p}_1^{\text{centre-specific}} = \frac{1}{n_c} \sum_{i=1}^{n_c} \text{expit} \left(\hat{\alpha} + \mathbf{L}'_i \hat{\beta} - \left(\frac{n_2}{n_1} \hat{\psi}_2 + \frac{n_3}{n_1} \hat{\psi}_3 + \dots + \frac{n_m}{n_1} \hat{\psi}_m \right) \right).$$

- A 'standardised odds ratio' can be estimated for each centre c as follows, if:

$$\hat{p}_{c,\text{overall}} = \frac{1}{n_c} \sum_{i=1}^{n_c} \text{expit}(\hat{\alpha} + \mathbf{L}'_i \hat{\beta}),$$

then

$$\text{OR}_{c,\text{overall}}^{\text{standardised}} = \frac{\hat{p}_c^{\text{centre-specific}} / (1 - \hat{p}_c^{\text{centre-specific}})}{\hat{p}_{c,\text{overall}} / (1 - \hat{p}_{c,\text{overall}})}.$$

¹This is the probability of 'success' that would have been observed had all patients in the study population been treated at centre c .

Discussion points

- The advantage of using weighted effect coding lies mainly in the interpretation of the intercept α and the fact that the 'mean' odds ratio is estimated differently for the standard effect coding (average centre risk):

$$\exp(\alpha) = \frac{(\prod_{c=1}^m \pi_c)^{1/m}}{(\prod_{c=1}^m (1 - \pi_c))^{1/m}},$$

and the weighted effect coding (average patient risk):

$$\exp(\alpha) = \frac{(\prod_{c=1}^m \pi_c^{n_c})^{1/n}}{(\prod_{c=1}^m (1 - \pi_c)^{n_c})^{1/n}},$$

with π_c the (adjusted) center-specific average risk.

In the second team meeting (16th of August, 2010) it was decided to focus on the *average of all centre-specific risks*, or hence standard effect coding.

- There are some problems when fitting these fixed-effect models when there are centres with only 'successes' or 'failures':
 - when there are many such extreme centres the maximum likelihood algorithm does not converge,
 - when there are few extreme centres, the standard errors corresponding to their estimate are huge,
 - when all extreme centres are discarded from the analysis, the problem is fixed.

A suggestion was to perform Firth's bias-reduced logistic regression (Firth (1993)); in R (version 2.10.1) the function for this (logistf) does not provide results (within less than 50 minutes) but in SAS (version 9.2) it works well. The standard errors of effects corresponding to extreme centres remain quite large, but the estimated effect is reasonable. Note that - even though this algorithm does not always converge - reasonable estimates can still be obtained. This appears to be a good alternative for the standard logistic regression on these data.

2.1.3 Hierarchical logistic regression model

A hierarchical logistic regression model with a patient- and centre level can be presented as follows:

- A patient-level (logistic regression model):

$$\text{logit}(E(Y_{ic}|\mathbf{L}_i)) = \alpha_c + \mathbf{L}_i'\boldsymbol{\beta}.$$

- A centre-level (linear regression model):

$$\alpha_c = \alpha + \psi_c, \quad \psi_c \sim N(0, \sigma_\psi^2)$$

Combining the two level model yields

$$\text{logit}(E(Y_{ic}|\mathbf{L}_i)) = \alpha + \mathbf{L}_i'\boldsymbol{\beta} + \psi_c, \quad \psi_c \sim N(0, \sigma_\psi^2).$$

Once the model is fitted using pseudo-likelihood estimation the *best linear unbiased predictors* (BLUP), $\hat{\psi}_c$, can be used to estimate the log-odds of 'success' in centre c versus the 'mean' log-odds of 'success' over all centres adjusting for the patient-specific covariates in the patient-level of the hierarchical model:

$$\text{OR}_{c,\text{overall}} = \exp(\hat{\psi}_c).$$

Each estimate $\hat{\psi}_c$ has an associated standard error $\text{SE}(\hat{\psi}_c)$ which is a function of the within-centre variance of centre c and the number of subjects in centre c relative to the other centres. Based on this, 95% confidence intervals for the estimated odds ratio's can be computed:

$$\left[\exp(\hat{\psi}_c - 1.96\widehat{\text{SE}}(\hat{\psi}_c)), \exp(\hat{\psi}_c + 1.96\widehat{\text{SE}}(\hat{\psi}_c)) \right],$$

where $\widehat{\text{SE}}(\hat{\psi}_c)$ can be obtained directly from standard software packages.

Extra outcome measures:

- The Procure-population averaged probability of 'success' can be estimated for each centre c as follows:

$$\hat{p}_c^{\text{pop.averaged}} = \frac{1}{n} \sum_{i=1}^n \text{expit}(\hat{\alpha} + \mathbf{L}_i'\hat{\boldsymbol{\beta}} + \hat{\psi}_c).$$

- The centre-specific probabilities of 'success' can be estimated for each centre c as follows:

$$\hat{p}_c^{\text{centre-specific}} = \frac{1}{n_c} \sum_{i=1}^{n_c} \text{expit}(\hat{\alpha} + \mathbf{L}_i'\hat{\boldsymbol{\beta}} + \hat{\psi}_c).$$

- A 'standardised odds ratio' can be estimated for each centre c as follows, if:

$$\hat{p}_{c,\text{overall}} = \frac{1}{n_c} \sum_{i=1}^{n_c} \text{expit}(\hat{\alpha} + \mathbf{L}_i'\hat{\boldsymbol{\beta}}),$$

then

$$\text{OR}_{c,\text{overall}}^{\text{standardised}} = \frac{\hat{p}_c^{\text{centre-specific}} / (1 - \hat{p}_c^{\text{centre-specific}})}{\hat{p}_{c,\text{overall}} / (1 - \hat{p}_{c,\text{overall}})}.$$

Discussion points

- Due to the shrinkage phenomenon centre-specific performances (e.g. odds ratios) are - in theory - closer to the mean (one) compared to the previous two methods.
- This model could be extended to a random intercepts- and slopes logistic regression model, i.e. that the patient-level effects would also be centre specific (β_c). Same comment for the centre-level of the model. For our primary analysis such interaction would not be considered. For the secondary analysis such random effect of patient-specific covariates may help explain the centre effect.

2.2 Survival outcomes

2.2.1 Cox' proportional hazards model

Call s the stratum 'cStage', then the stratified Cox' proportional hazards model is defined for patient i in stratum s as:

$$\lambda_i(t) = \lambda_{0,s_i}(t) \exp(\mathbf{L}_i' \boldsymbol{\beta} + \sum_{c=2}^m \psi_c X_{ic}), \quad (5)$$

where $\lambda_0(t)$ is the unspecified baseline hazard function and X_{ic} ($c = 2, \dots, m$) are standard (un-weighted) effect coded variables for the centre choice.

The probability of 5-year survival for patient i in stratum s_i in centre c can be estimated as follows (for $c \neq 1$):

$$\hat{p}_{i,c}^5 = P(T_i > 5 | S_i, \mathbf{L}_i, \mathbf{X}_i = c) = \hat{S}_{0,s_i}(5)^{\exp(\mathbf{L}_i' \hat{\boldsymbol{\beta}} + \hat{\psi}_c)},$$

and for $c = 1$:

$$\hat{p}_{i,c}^5 = P(T_i > 5 | S_i, \mathbf{L}_i, \mathbf{X}_i = 1) = \hat{S}_{0,s_i}(5)^{\exp(\mathbf{L}_i' \hat{\boldsymbol{\beta}} - (\hat{\psi}_2 + \hat{\psi}_3 + \dots + \hat{\psi}_m))},$$

with

$$\hat{S}_{0,s}(5) = \exp\left(-\int_0^5 \hat{\lambda}_{0,s}(u) du\right),$$

where the cumulative stratum-specific baseline hazard corresponds to the standard Breslow estimate.

Outcome measures:

- Centre-specific 5-year survival is then obtained by taking the average of the estimated 5-year survival probabilities of all patients within the same centre:

$$\hat{p}_c^{5,\text{centre-specific}} = \frac{1}{n_c} \sum_{i=1}^{n_c} \hat{p}_{i,c}^5.$$

- The Procure-population averaged 5-year survival probabilities can be estimated for each centre c as follows (for $c \neq 1$):

$$\hat{p}_c^{5,\text{pop.averaged}} = \frac{1}{n} \sum_{i=1}^n \hat{S}_{0,s_i}(5)^{\exp(\mathbf{L}_i' \hat{\boldsymbol{\beta}} + \hat{\psi}_c)},$$

and for $c = 1$:

$$\hat{p}_1^{5,\text{pop.averaged}} = \frac{1}{n} \sum_{i=1}^n \hat{S}_{0,s_i}(5)^{\exp(\mathbf{L}_i' \hat{\boldsymbol{\beta}} - (\hat{\psi}_2 + \hat{\psi}_3 + \dots + \hat{\psi}_m))}.$$

- A 'standardised odds ratio' can be estimated for each centre c as follows, if:

$$\hat{p}_{c,\text{overall}}^5 = \frac{1}{n_c} \sum_{i=1}^{n_c} \hat{S}_{0,s_i}(5)^{\exp(\mathbf{L}_i' \hat{\boldsymbol{\beta}})},$$

then

$$\text{OR}_{c,\text{overall}}^{\text{standardised}} = \frac{\hat{p}_c^{5,\text{centre-specific}} / (1 - \hat{p}_c^{5,\text{centre-specific}})}{\hat{p}_{c,\text{overall}}^5 / (1 - \hat{p}_{c,\text{overall}}^5)}.$$

Discussion point

There are some problems when fitting these Cox proportional hazards effect models when there are centres without events:

- when there are few extreme centres, the standard errors corresponding to their estimate are huge,
- when all extreme centres are discarded from the analysis, the problem is fixed.

A suggestion was to perform Firth's bias-reduced maximum likelihood estimation (Firth 1993); in R (version 2.10.1) this function (`coxphf`) does not provide results but in SAS (version 9.2) it works well. The standard errors of effects corresponding to extreme centres remain quite large, but the estimated effect is reasonable. Note that reasonable estimates can still be obtained as confirmed by our simulations. This appears to be a good alternative for the standard Cox proportional hazards modelling on these data.

2.2.2 Cox' frailty proportional hazards model

Frailty models are frequently used for modelling dependence in time-to-event data. They are the Cox-model equivalent of the random effects model. The aim of the frailty is to take the presence of correlation - due to some shared covariate information - between survival times into account, and correct for it without needing to fit a separate hazard (intercept) for each centre through a parametric model.

We will assume a *constant shared frailty*, or all patients i in centre c share the same frailty ψ_c . The variability of the ψ_c 's reflects the heterogeneity of risks between the m centres.

The conditional hazards model for patient i in stratum s_i in centre c is an extension of model (5)

$$\lambda_i(t, \psi) = \lambda_{0,s_i}(t) \exp(\mathbf{L}_i' \boldsymbol{\beta} + \psi_c) \quad (6)$$

The model assumes that all observed event-times are independent given the frailties, hence assumed a 'conditional independence' model. The value of each component of $\boldsymbol{\psi} = (\psi_1, \dots, \psi_m)^t$ is constant over time and common to the patients in the centres (across all strata) and thus responsible for creating dependence within centre c .

In practice, the frailties $\exp(\psi_c)$ are assumed to be independent and identically distributed with mean $E[\exp(\Psi)] = 1$, and some unknown variance $\text{Var}[\exp(\Psi)] = \sigma_\psi^2$. For computational convenience and convergence, the frailty distribution is often taken to be a gamma distribution $\exp(\Psi) \sim \text{Gamma}(k, \theta)^2$. For identifiability reasons we suppose here that $k = \theta$, yielding a random effect in model (6) with

$$\exp(\Psi) \sim \text{Gamma}(1/\sigma_\psi^2, 1/\sigma_\psi^2), \quad (7)$$

see e.g. Figure 14.

The probability of 5-year survival for patient i in stratum s in centre c can be estimated as follows:

$$\hat{p}_{i,c}^5 = P(T_i > 5 | S_i, \mathbf{L}_i, \mathbf{X}_i = c) = \hat{S}_{0,s_i}(5)^{\exp(\mathbf{L}_i' \hat{\boldsymbol{\beta}} + \hat{\psi}_c)},$$

with

$$\hat{S}_{0,s}(5) = \exp\left(-\int_0^5 \hat{\lambda}_{0,s}(u) du\right),$$

where the cumulative stratum-specific baseline hazard corresponds to the standard Breslow estimate.

Outcome measures:

- Centre-specific 5-year survival is then obtained by taking the average of the estimated 5-year survival probabilities over all patients within the same centre:

$$\hat{p}_c^{5, \text{centre-specific}} = \frac{1}{n_c} \sum_{i=1}^{n_c} \hat{p}_{i,c}^5.$$

²The probability density function of a random variable $X \sim \text{Gamma}(k, \theta)$, with θ a *scale* parameter and k a *shape* parameter is defined as:

$$f_X(x) = x^{k-1} \theta^k \frac{e^{-x\theta}}{\Gamma(k)},$$

with $x, k, \theta > 0$.

- The Procure-population averaged 5-year survival probabilities can be estimated for each centre c as follows:

$$\hat{p}_c^{5, \text{pop. averaged}} = \frac{1}{n} \sum_{i=1}^n \hat{S}_{0, s_i}(5)^{\exp(\mathbf{L}_i' \hat{\beta} + \hat{\psi}_c)}.$$

- A 'standardised odds ratio' can be estimated for each centre c as follows, if:

$$\hat{p}_{c, \text{overall}}^5 = \frac{1}{n_c} \sum_{i=1}^{n_c} \hat{S}_{0, s_i}(5)^{\exp(\mathbf{L}_i' \hat{\beta})},$$

then

$$\text{OR}_{c, \text{overall}}^{\text{standardised}} = \frac{\hat{p}_c^{5, \text{centre-specific}} / (1 - \hat{p}_c^{5, \text{centre-specific}})}{\hat{p}_{c, \text{overall}}^5 / (1 - \hat{p}_{c, \text{overall}}^5)}.$$

3 Propensity scores

3.1 Binary outcomes

Focus is on the probability of 'success' that would have been observed had all patients in the study population been treated at centre c . Let $Y_i(c)$ denote the probability of 'success' that would have been observed for a given patient i if treated at centre c . Then more formally our interest lies in

$$P\{Y(c) = 1\} = E\{Y_i(c)\}.$$

The outcome working model is a logistic regression model for the probability of 'success' in centre c in function of patient-specific characteristics, i.e.

$$m(c, \mathbf{L}; \alpha_c, \beta_c) = \text{logit}(E(Y_i | \mathbf{X}_i = c, \mathbf{L}_i)) = \alpha_c + \mathbf{L}'_i \beta_c,$$

for each centre c , or this may also be a joint model for all centres (as in (4)):

$$m(\mathbf{L}; \alpha, \beta, \psi) = \text{logit}(E(Y_i | \mathbf{X}_i, \mathbf{L}_i)) = \alpha + \mathbf{L}'_i \beta + \sum_{c=2}^m \psi_c X_{ic}.$$

The propensity score working model is a multinomial regression model for the multiple propensity score: the probability of a patient being treated in center c in function of patient-specific characteristics, i.e.

$$h(c, \mathbf{L}; \alpha_c^*, \beta_c^*) = P(I(\mathbf{X}_i = c) | \mathbf{L}_i) = \begin{cases} \frac{1}{1 + \sum_{j=2}^m \exp(\alpha_j^* + \mathbf{L}'_j \beta_j^*)} & c = 1 \\ \frac{\exp(\alpha_c^* + \mathbf{L}'_c \beta_c^*)}{1 + \sum_{j=2}^m \exp(\alpha_j^* + \mathbf{L}'_j \beta_j^*)} & c \neq 1 \end{cases}.$$

Given estimators of the parameters in both working models, we propose to estimate $E\{Y(c)\}$ as

$$E\{\widehat{Y(c)}\} = \frac{1}{n} \sum_{i=1}^n m(\mathbf{L}_i; \hat{\alpha}, \hat{\beta}, \hat{\psi}_c) + \frac{1}{n} \sum_{i=1}^n \frac{X_{ic}}{h(c, \mathbf{L}_i; \hat{\alpha}_c^*, \hat{\beta}_c^*)} \left\{ Y_i - m(\mathbf{L}_i; \hat{\alpha}, \hat{\beta}, \hat{\psi}_c) \right\}.$$

It can be shown using similar arguments as in Tsiatis et al. (2008) that this estimator is doubly robust in the sense that it is a unbiased estimator of $E\{Y(c)\}$ (in sufficiently large samples) if either the outcome regression model or the propensity score model is correctly specified.

The above doubly robust estimator may give probability estimates outside the $[0 - 1]$ interval. A first modification which guarantees better performance is to use the following estimator

$$E\{\widehat{Y(c)}\} = \frac{1}{n} \sum_{i=1}^n m(\mathbf{L}_i; \hat{\alpha}, \hat{\beta}, \hat{\psi}_c) + \frac{\sum_{i=1}^n \frac{X_{ic}}{h(c, \mathbf{L}_i; \hat{\alpha}_c^*, \hat{\beta}_c^*)} \left\{ Y_i - m(\mathbf{L}_i; \hat{\alpha}, \hat{\beta}, \hat{\psi}_c) \right\}}{\sum_{i=1}^n \frac{X_{ic}}{h(c, \mathbf{L}_i; \hat{\alpha}_c^*, \hat{\beta}_c^*)}}.$$

By dividing by the sum of the weights in the second term, we guarantee a much better behavior in small samples. The resulting estimator retains the double robustness property as can be demonstrated using similar arguments as in Goetgeluk et al. (2008).

A second modification is to fit the outcome working model via a weighted regression of outcome on covariates amongst patients attending center c , with weights $1/h(c, \mathbf{L}_i; \hat{\alpha}_c^*, \hat{\beta}_c^*)$ (Robins et al. (2008)). When the outcome regression includes an intercept and is fitted using standard software, then the implication is to set

$$\frac{1}{n} \sum_{i=1}^n \frac{X_{ic}}{h(c, \mathbf{L}_i; \hat{\alpha}_c^*, \hat{\beta}_c^*)} \left\{ Y_i - m(c, \mathbf{L}_i; \hat{\alpha}, \hat{\beta}, \hat{\psi}_c) \right\} = 0$$

so that the doubly robust estimator can simply be obtained as

$$E\{\widehat{Y(c)}\} = \frac{1}{n} \sum_{i=1}^n m(c, \mathbf{L}_i; \widehat{\alpha}, \widehat{\beta}, \widehat{\psi}_c).$$

Apart from simplicity, an attraction of this estimator is that, unlike the previous doubly robust estimator, it does not inflate the bias due to model misspecification in regions where the weights are large (Vansteelandt et al. (2010)). A further attraction is that, because the estimator is calculated as the average of the fitted values from a binary outcome regression model, it guarantees probability estimates within the $[0 - 1]$ interval. We will refer to this estimator as the 'regression doubly robust estimator'.

To ensure that the weights $1/h(c, \mathbf{L}_i; \widehat{\alpha}_c^*, \widehat{\beta}_c^*)$ in the analysis are not too variable, we have followed a recommendation by Cole and Hernan (2008) in all analyses to truncate all weights at the 1%- and 99%-tile. This implies that weights larger than the 99%-tile are set to the 99%-tile and weights smaller than the 1%-tile are set to the 1%-tile.

Discussion points

- Similarly as for the fixed-effects regression method, the Firth correction could be used in the outcome working model.
- Because centre is a discrete covariate with many levels, one may also consider shrinkage estimators for the centre effects in the outcome working model as may be obtained from random effects models. This is likely to be promising, but was not considered in this report as, to the best of our knowledge, it has not been studied with propensity scores in the literature.
- In theory, further improvements may be attained by fitting also the propensity score model via a weighted regression of center on covariates with weights $1/h(c, \mathbf{L}_i; \widehat{\alpha}_c^*, \widehat{\beta}_c^*)$ (Vansteelandt et al. (2010)). We have chosen not to pursue this strategy because the number of subjects per center was too small to guarantee well-performing propensity score estimates in this way.
- Standard errors of the proposed estimators of $E\{Y(c)\}$ can be obtained via sandwich estimators.

3.2 Survival outcomes

With a survival outcome, the focus will be on the t -year survival probability $S(t) = P(Y > t)$. If there were no censoring, then inference for $S(t)$ would follow the lines described in the previous section, but in the presence of censoring, we will rely on inverse probability of censoring weighting to make progress. Assume that L (i.e. the previously considered sufficient set of covariates to adjust for differential patient mix) also contains all predictors of censoring before time t that are associated with survival up to time t , so that with C and T the censoring time and survival time, respectively,

$$I(C > t) \perp\!\!\!\perp I(T > t) | X, L.$$

The inverse probability of censoring weighted estimators that we develop, rely on a censoring working model which is a logistic regression model for the probability of censoring after t years

$$d(\mathbf{X}, \mathbf{L}; \alpha^\dagger, \beta^\dagger, \psi^\dagger) = \text{logit}(P(C_i > t | \mathbf{X}_i, \mathbf{L}_i)) = \alpha^\dagger + \mathbf{L}_i' \beta^\dagger + \sum_{c=2}^m \psi_c^\dagger X_{ic}.$$

The outcome working model is a hazard regression model in all centres in function of patient-specific characteristics, i.e.

$$g(t, \mathbf{L}; \beta, \psi) = \lim_{\Delta t \rightarrow 0} \frac{f(T_i < t + \Delta t | T_i \geq t, \mathbf{X}_i, \mathbf{L}_i)}{\Delta t} = \lambda_{0, s_i}(t) \exp(\mathbf{L}_i' \beta + \sum_{c=2}^m \psi_c X_{ic}),$$

from which

$$m(t, L; \beta, \psi) = P(T_i > t | \mathbf{X}_i, \mathbf{L}_i) = \exp \left\{ - \int_0^t \lambda_{0, s_i}(u) \exp(\mathbf{L}_i' \beta + \sum_{c=2}^m \psi_c X_{ic}) du \right\}.$$

The reason for choosing a hazard regression model for the outcome is because standard inference for such models naturally accommodates censoring under the above considered censoring assumption.

We now propose to estimate the counterfactual survival time corresponding to center c , as

$$\begin{aligned} \hat{P}\{Y(c) > t\} &= \frac{1}{n} \sum_{i=1}^n m(\mathbf{L}_i; \hat{\beta}, \hat{\psi}_c) + \\ &\frac{1}{n} \sum_{i=1}^n \frac{I(C_i > t)}{d(\mathbf{X}_i, \mathbf{L}_i; \hat{\alpha}^\dagger, \hat{\beta}^\dagger, \hat{\psi}_c^\dagger)} \frac{X_{ic}}{h(c, \mathbf{L}_i; \hat{\alpha}_c^*, \hat{\beta}_c^*)} \left\{ I(T_i > t) - m(\mathbf{L}_i; \hat{\beta}, \hat{\psi}_c) \right\}. \end{aligned}$$

As in the previous section, truncation of the weights at the 1%- and 99%-tile (Cole and Hernan (2008))

$$\frac{I(C_i > t)}{d(\mathbf{X}_i, \mathbf{L}_i; \hat{\alpha}^\dagger, \hat{\beta}^\dagger, \hat{\psi}_c^\dagger)} \frac{X_{ic}}{h(c, \mathbf{L}_i; \hat{\alpha}_c^*, \hat{\beta}_c^*)}$$

is possible to ensure that they are not too variable.

It can be shown that this estimator is doubly robust in the sense that it is an unbiased estimator of $S(t, c)$ (in large samples) if either the outcome regression model is correctly specified, or both the censoring model and the propensity score model are correctly specified.

Discussion points

- We do not consider inference under the weaker assumption that

$$I(C > t) \perp\!\!\!\perp I(D > t) | X, W,$$

where $L \subseteq W$. We have chosen not to do so because, in the available data, we have no access to additional potential predictors of censoring over and above those already contained in L . For the same reason, we have also chosen not to allow for time-varying predictors of censoring. However, the formalism that we develop relatively easily extends to enable these relaxations.

- Similarly as for the fixed-effects regression method, the Firth correction could be used in the outcome working model and a hierarchical version could be considered to make methods comparable with the regression methods.
- Standard errors of the proposed estimators can be obtained via sandwich estimators.

4 Simulations based on the original Procure database

To investigate the practical performance in the Procure setting of the methods described above, simulations based on the Procure database are performed on two selected QCI's:

- QCI 1111: Overall 5-year survival by stage (right-censored survival outcome)
- QCI 1232a: Proportion of APR and Hartmann's procedures (binary outcome with known variation between the centres)

The QCI's were selected as representative of the different types of outcomes.

Currently, the Procure database contains $n = 2901$ patients. The following three patient-specific confounders were selected for risk-adjustment in the simulations:

- Gender: 61% males and 39% females (no missings)
- Age (116 missing)
 - mean = 67.32
 - standard deviation = 11.7

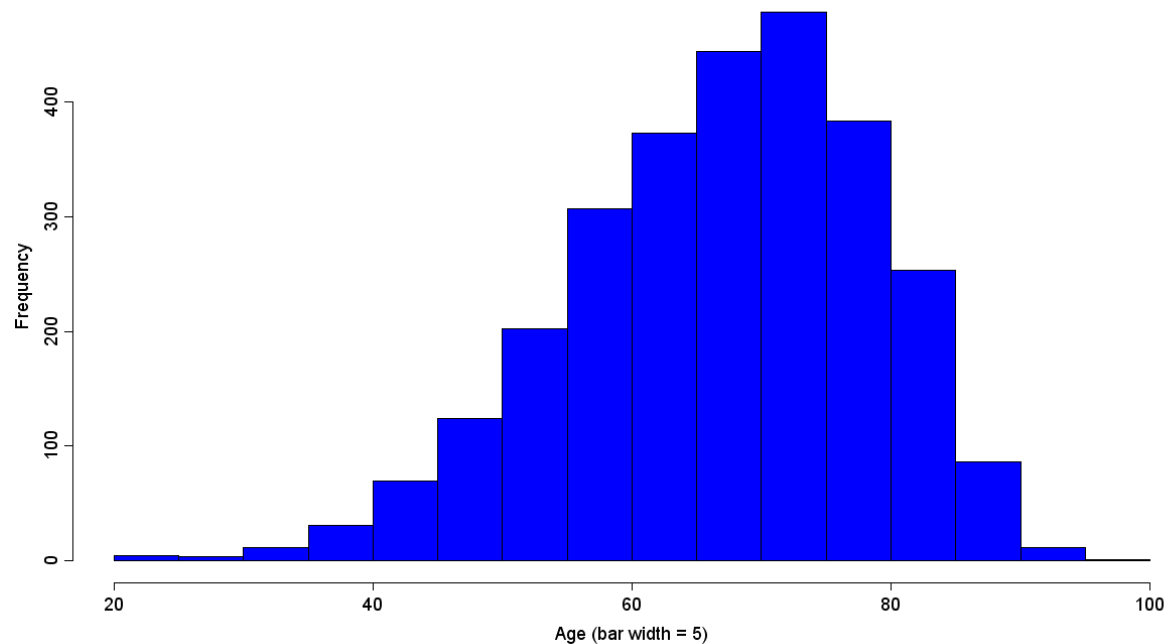


Figure 1: Distribution of age in the total Procure population

For all patients with missing age, the incidence date³ was missing.

³This is the date of biopsy, the date of first consultation or the date of first treatment. If none of these was given, the incidence date is missing, but age could in principle still be approximated based on the earliest date available.

- cStage

0	I	II	III	IV	X	missing
12	341	459	1323	357	98	311
0.4%	11.8%	15.8%	45.6%	12.3%	3.4%	10.7%

Patients with cStage 0 will be discarded from all further analyses. Missingness in cStage is partly centre-specific (Figure 2) and tends to occur more frequently in small centres than in large centres.

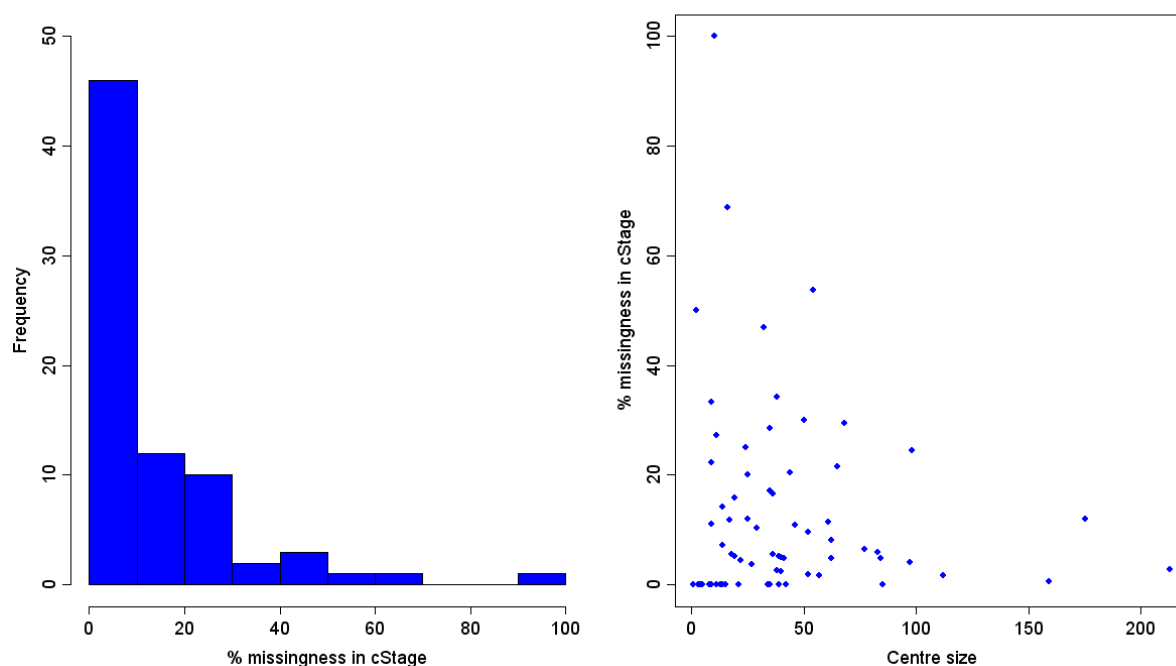


Figure 2: Distribution of % missingness in cStage over the centres and the relation between the centre size and % missingness

It appears that 115 of the patients with missing age also have missing cStage (one patient with missing age has cStage = I). Further investigation of these patients might reveal that most were registered early, since the old version of the CRF has no explicit biopsy date.

4.1 Simulation protocol

Patients with missing values for one of the selected patient-specific confounders (or cStage = 0) are automatically discarded from the simulations, leaving us with $n = 2577$ patients for the simulation exercise. The selected patients came from 75 centres with a varying range of sizes (Figure 3).

Clearly there are a substantial number of (very) small centres (Table 1) for which statistical analysis becomes challenging with necessarily imprecise results. As agreed upon in the second UGent-team meeting (August 16th, 2010), centres with less than 5 patients are *grouped* in one overlapping centre for all statistical analyses.

Note that the actual number of patients (n) and centres (m) considered might differ between the considered QCI's at they might have their own - sometimes structural - missingness patterns, to be

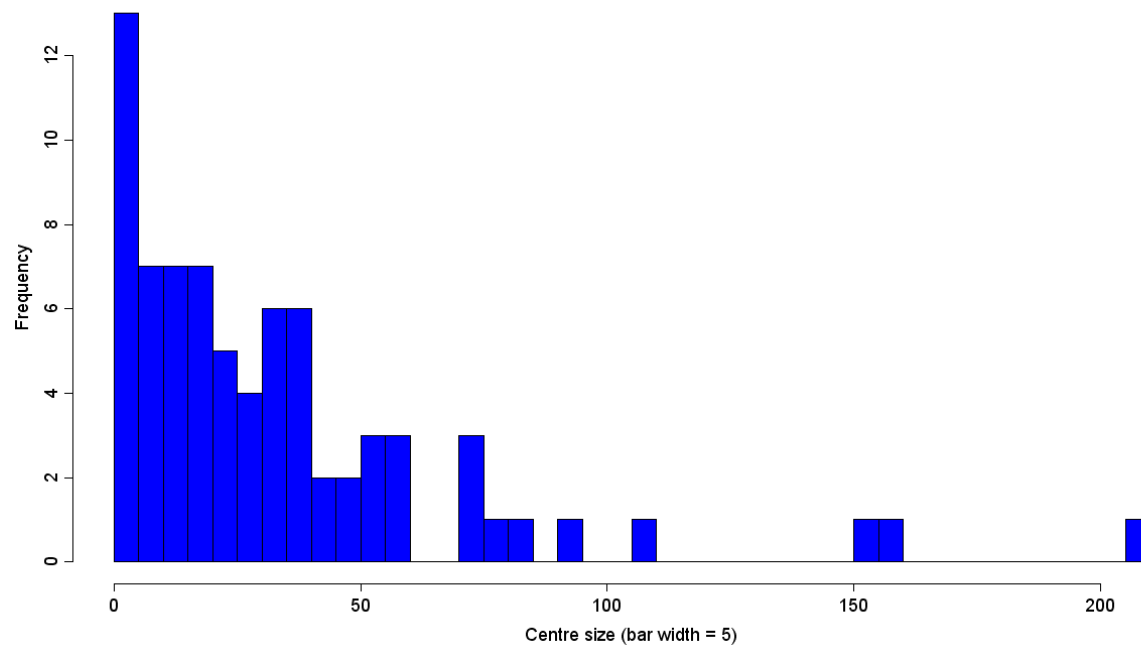


Figure 3: Distribution of centre sizes in the Procure population selected for the simulations

# patients	Frequency	Cumulative Frequency	Cumulative %
1	6	6	0.08
3	2	8	0.11
4	2	10	0.13
5	3	13	0.17
6	1	14	0.19
7	2	16	0.21
8	3	19	0.25
9	1	20	0.27
11	1	21	0.28

Table 1: Distribution of centres sizes in the Procure population selected for the simulations

discussed further on.

Table 2 gives an overview of the outcome measures (per centre) that will be obtained from each simulated dataset.

Type QCI	Statistical analysis	Outcome measure
Binary	O/E method	Standardised event rate (SER)
	Fixed-effect logistic regression(*)	Odds ratio
		Centre-specific probability of 'success' Population-averaged probability of 'success' (Centre-specific) standardised odds ratio
	Hierarchical logistic regression	Odds ratio
		Centre-specific probability of 'success' Population-averaged probability of 'success' (Centre-specific) standardised odds ratio
	Propensity score: regression doubly robust estimator	Population-averaged probability of 'success'
Survival	Cox proportional hazards model(*)	Centre-specific probability of x-year survival Population-averaged probability of x-year survival' (Centre-specific) standardized odds ratio
		Centre-specific probability of x-year survival Population-averaged probability of x-year survival (Centre-specific) standardised odds ratio
	Frailty proportional hazards model	Centre-specific probability of x-year survival Population-averaged probability of x-year survival (Centre-specific) standardised odds ratio
		Population-averaged probability of x-year survival
	Propensity score: doubly robust estimator	Population-averaged probability of x-year survival

(*): Firth's bias reduced maximum likelihood estimation method is used instead of standard maximum likelihood estimation (Firth (1993), Heinze and Schemper (2002) and Heinze and Schemper (2001))

Table 2: Outcome measures (per centre) to be estimated in each simulated dataset

The 'true' outcome models are different depending on the type of QCI, and therefore also the strategy for simulating outcomes for a given set of patient-specific confounders and centre choice:

- Binary: A hierarchical logistic regression model was first fitted to the original data. Its estimated parameters were fixed and from the estimated distribution of random centre effects one set of new centre effects were generated and fixed. Using all these parameters, independent simulated datasets were generated based on a fixed effects model.
- Survival: A frailty proportional hazards model was first fitted to the original data. Its estimated baseline hazard and regression parameters were fixed and from the estimated distribution of frailty centre effects one set of new centre effects were generated and fixed. Using all these parameters, independent simulated datasets were generated based on a fixed effects model.

The outcomes of interest in Table 2 have been estimated for each centre in each simulated dataset (if analytically practical together with a 95% confidence interval for this outcome). Using all this information it is possible to compare the different methods based on the following summary measures of the simulated outcomes measures:

- per centre, the 2.5%- and 97.5%-tile of the estimated outcomes over all simulated datasets are used to determine respectively the lower and upper limit of the *95% empirical confidence interval*,
- per centre, the *coverage* of the 95% empirical confidence interval is obtained as

$$\text{coverage}_c^x = \frac{1}{1000} \sum_{b=1}^{1000} I(\hat{x}_c^b - w_c^L < x_c < \hat{x}_c^b + w_c^U), \quad (8)$$

with x_c the 'true' outcome measure of interest for centre c and \hat{x}_c^b the outcome measure of interest estimated for centre c in the b th simulated dataset, and

$$w_c^L = \text{median}_b(\hat{x}_c^b) - \text{LL}_c^x,$$

$$w_c^U = UL_c^x - \text{median}_b(\hat{x}_c^b),$$

with LL_c^x and UL_c^x respectively the lower and upper limit of the 95% empirical confidence interval for the outcome measure x .

Over the m centres, the estimated coverages for a specific outcome measure are statistically summarized by the minimum, 25%-tile, median, 75%-tile and the maximum.

- over all centres the *median width* of these 95% empirical confidence intervals is obtained,
- per centre, the *root mean squared error* (RMSE) is obtained as

$$RMSE_c^x = \sqrt{\frac{1}{1000} \sum_{b=1}^{1000} (\hat{x}_c^b - x_c)^2}, \quad (9)$$

with x_c the 'true' outcome measure of interest for centre c and \hat{x}_c^b the outcome measure of interest estimated for centre c in the b th simulated dataset.

Over the m centres, the estimated RMSE's for a specific outcome measure are statistically summarized by the minimum, 25%-tile, median, 75%-tile and the maximum.

Table 3 gives an overview of the software packages and specific functions/procedures used for the respective methods used in the simulation exercise.

Type QCI	Method	Software package	Function/procedure
Binary	O/E method	R (2.10.1)	lrm {Design}
	Fixed-effect logistic regression	SAS (9.2)	PROC LOGISTIC with firth-option in model-statement
	Hierarchical logistic regression	R (2.10.1)	lmer {lme4}
	Propensity score	R (2.10.1)	glm {stats} multinom {nnet}
Survival	Cox proportional hazards model	SAS (9.2)	PROC PHREG with firth-option in model-statement
	Frailty proportional hazards model	R (2.10.1)	coxph {survival}
	Propensity score	SAS (9.2)	PROC PHREG with firth-option in model-statement
		R (2.10.1)	multinom {nnet}

Table 3: Software packages (version) and specific functions/procedures used in the simulation exercise

4.1.1 Version 1: Use correct random effects distribution

Preparatory steps

1. A propensity score model - relating the centre choice to selected patient-specific confounders - was fitted using a multinomial logistic regression model with main effects for age, gender and cStage. This yields a $(n \times m)$ -matrix \widehat{PS} with estimated probabilities of patient $i = 1, \dots, n$ being treated in centre $c = 1, \dots, m$.
2. The 'true' outcome model with as predictors the main effects for age, gender and cStage is fitted on the original dataset (with the grouped small centres) and its estimated parameters $(\widehat{\alpha}, \widehat{\beta}, \widehat{\sigma}_{\psi})$ are considered as the true (population) parameters.

Actual simulations Using the estimated parameters of the 'true' model $(\widehat{\alpha}, \widehat{\beta}, \widehat{\sigma}_{\psi})$ and estimated probabilities of being treated in each centre (\widehat{PS}), 1000 simulated datasets (with new centre choice and outcome) are created from a multinomial distribution for the centre choice and a fixed effects model for the outcome (cfr. p 23).

4.1.2 Version 2: Incorporate misspecification of the random effects distribution

Preparatory steps

1. A propensity score model - relating the centre choice to the selected patient-specific confounders - was fitted as before using a multinomial logistic regression model with main effects for age, gender and cStage. This yields a $(n \times m)$ -matrix \widehat{PS} with estimated probabilities of patient $i = 1, \dots, n$ being treated in centre $c = 1, \dots, m$.
2. A 'true' outcome model with as predictors the main effects age, gender and cStage is fitted on the original dataset (with the grouped small centres) and estimated parameters $(\widehat{\alpha}, \widehat{\beta}, \widehat{\sigma}_{\psi})$ are considered as the true (population) parameters.
3. To propose a realistic random effects distribution which deviates from the global normal variation in random effects, one (or more) breakpoints (N_c) in terms of number of patients per centre is chosen in order to define small versus large centres. An 'optimal' breakpoint(s) N_c is determined such that the outcome measure is most different between the small and large centres.
4. Using the 'optimal' N_c breakpoint a bimodal (or multimodal) distribution for the random effects will be constructed, and 'true' random effects ψ_c^{true} will be generated from this distribution based on $\widehat{\sigma}_{\psi}$ from point 2.

Actual simulations Using the estimated parameters of the 'true' model $(\widehat{\alpha}, \widehat{\beta}, \psi_c^{\text{true}})$, the optimal breakpoint(s) N_c and estimated probabilities of being treated in each centre (\widehat{PS}), 1000 simulated datasets (with new centre choice and outcome) are created from a multinomial distribution for the centre choice and a fixed effects model for the outcome (cfr. p 23).

4.2 Simulation results for the binary outcome: Proportion of APR and Hartmanns procedures (QCI 1232a)

This QCI is defined in Vlayen et al. (2008) as:

- Numerator: all patients with RC that underwent radical resection and had an APR or Hartmanns procedure.
- Denominator: all patients with RC that underwent radical resection.

There are $n = 2355$ patients with full information on the patient-specific confounders that underwent radical resection. After grouping small centres (with less than 5 patients meeting these criteria) into one overlapping centre, $m = 64$ centres are available for the simulation exercise (note that centre $c = 64$ is the overlapping centre). The propensity scores \widehat{PS} are estimated from a multinomial regression model with as covariates the main effects age, gender and cStage (Figure 4).

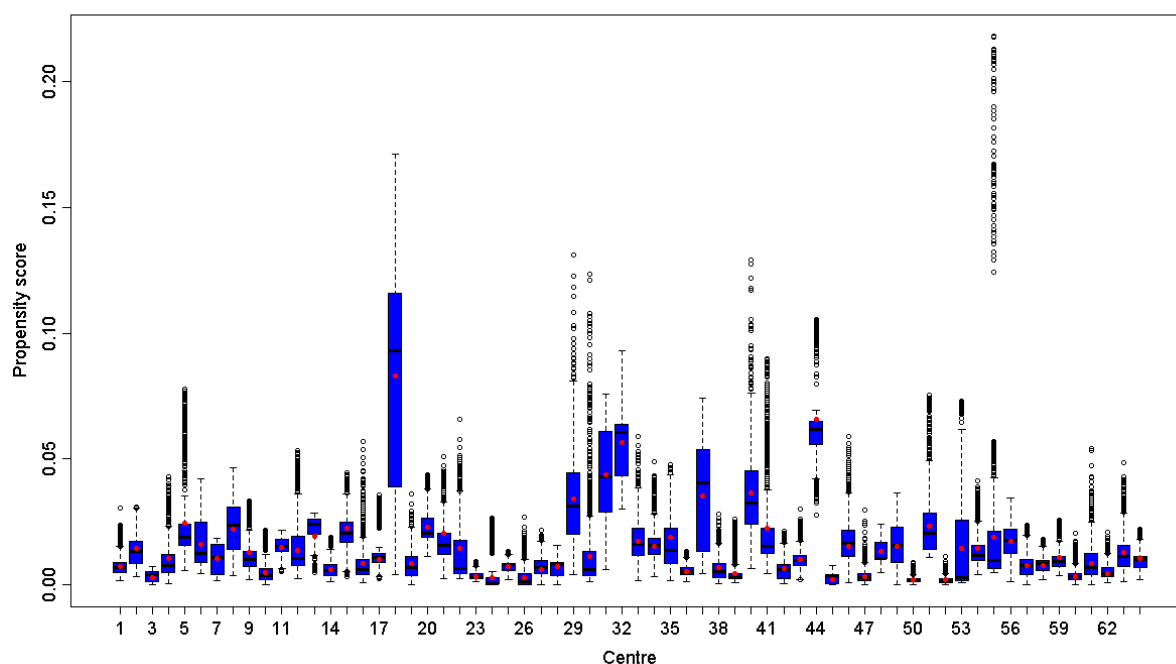


Figure 4: Distribution of estimated propensity scores per centre

Overall, 23.9% of the patients selected because they underwent radical resection had an APR or Hartmanns procedure (we call this a 'success'). The distribution of the outcome over the cStages is given in Table 4.

The number of patients and 'successes' per centre have been checked and no anomalies were seen [not shown to protect confidentiality].

The 'true' outcome model is the hierarchical logistic regression model with as patient-specific covariates L the main effects age, gender and cStage:

$$\text{logit}(P(Y_{ic} = 1 | \mathbf{L}_i)) = \alpha + \mathbf{L}'_{ic}\boldsymbol{\beta} + \psi_c,$$

	Y = 0	Y = 1	% success
I	270	48	15.1%
II	309	125	28.8%
III	960	300	23.8%
IV	191	73	27.7%
X	61	18	22.8%

Table 4: Distribution of outcome per cStage

with $\Psi_c \sim \text{Normal}(0, \sigma_\Psi^2)$. The estimated parameters $(\hat{\alpha}, \hat{\beta}, \hat{\sigma}_\Psi^2)$ are given in the R-output below.

Generalized linear mixed model fit by the Laplace approximation

Formula: Y ~ L + (1 | X.orig)

AIC BIC logLik deviance
2534 2580 -1259 2518

Random effects:

Groups Name Variance Std.Dev.
X.orig (Intercept) 0.18697 0.4324

Number of obs: 2355, groups: X.orig, 64

Fixed effects:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-1.157286	0.315934	-3.663	0.000249 ***
Lgender	-0.028629	0.102635	-0.279	0.780295
Lage	0.020221	0.004553	4.441	8.94e-06 ***
Lstage1	-0.576541	0.322828	-1.786	0.074114 .
Lstage2	0.245348	0.297788	0.824	0.409996
Lstage3	0.076737	0.287358	0.267	0.789435
Lstage4	0.265878	0.311947	0.852	0.394037

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Direct comparisons in performance between the different methods can already be made on the original data, namely by comparing the distribution of the estimated outcomes measures (the centre-specific and population averaged probability of 'success' and the centre-specific standardised odds ratio) between the different methods applied (Firth corrected fixed-effects logistic regression, hierarchical logistic regression and the propensity score method), see Figures 5 and 6.

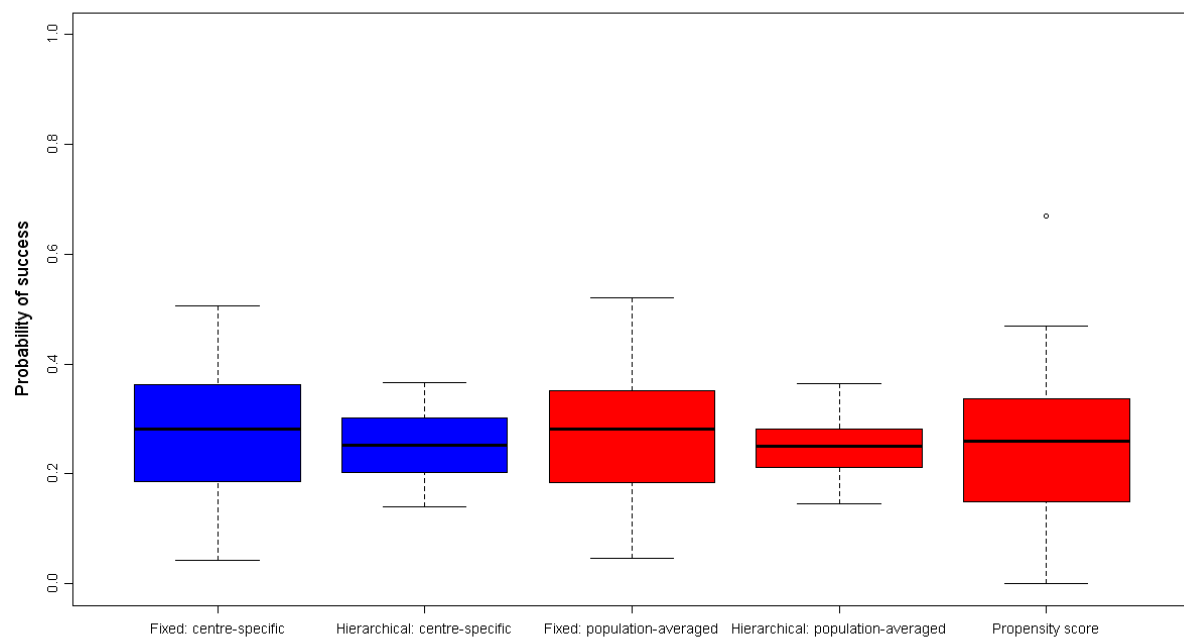


Figure 5: Comparison of centre-specific and population averaged probabilities of success for QCI 1232a for the three applied statistical methods. The probabilities in this graph are estimated from the original Procure dataset.

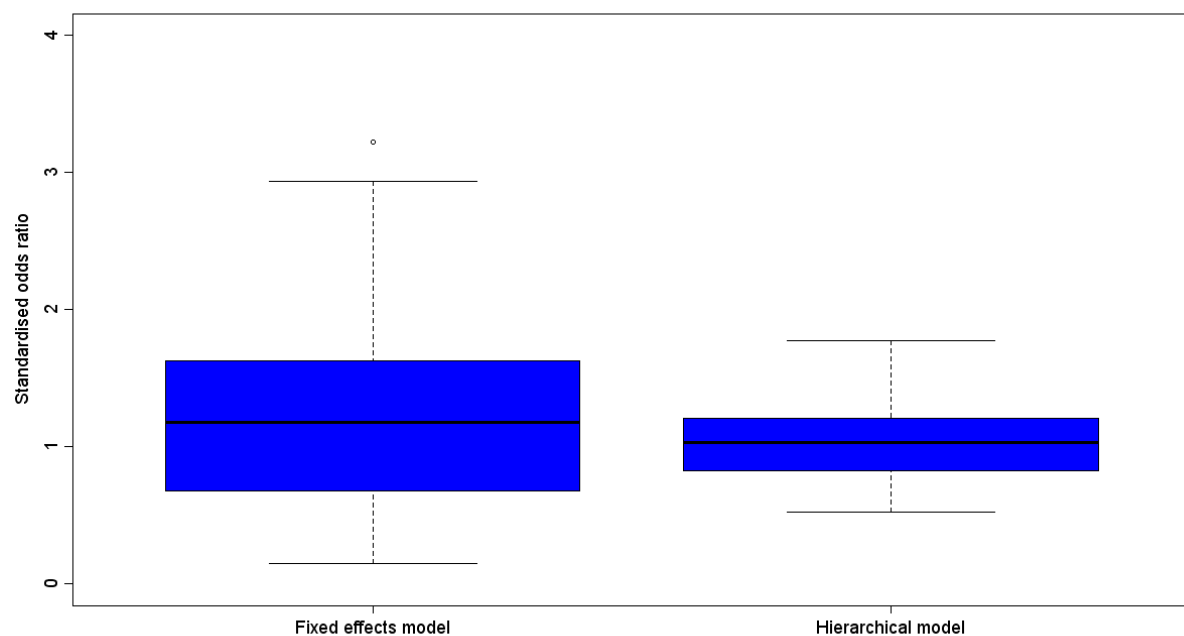


Figure 6: Comparison of the (centre-specific) standardised odds ratio of success for QCI 1232a between the (Firth corrected) fixed-effects logistic regression and hierarchical logistic regression methods. The odds ratio's in this graph are estimated from the original Procure dataset.

4.2.1 Version 1: Use correct random effects distribution

Using the $\hat{\sigma}_\psi = 0.43$ the 'true' centre effects are generated as follows

$$\psi_c^{\text{true}} \sim \text{Normal}(0, \hat{\sigma}_\psi^2).$$

The distribution of the BLUPs from the 'true' outcome model and of the generated 'true' centre effects (ψ_c^{true}) are shown in Figure 7.

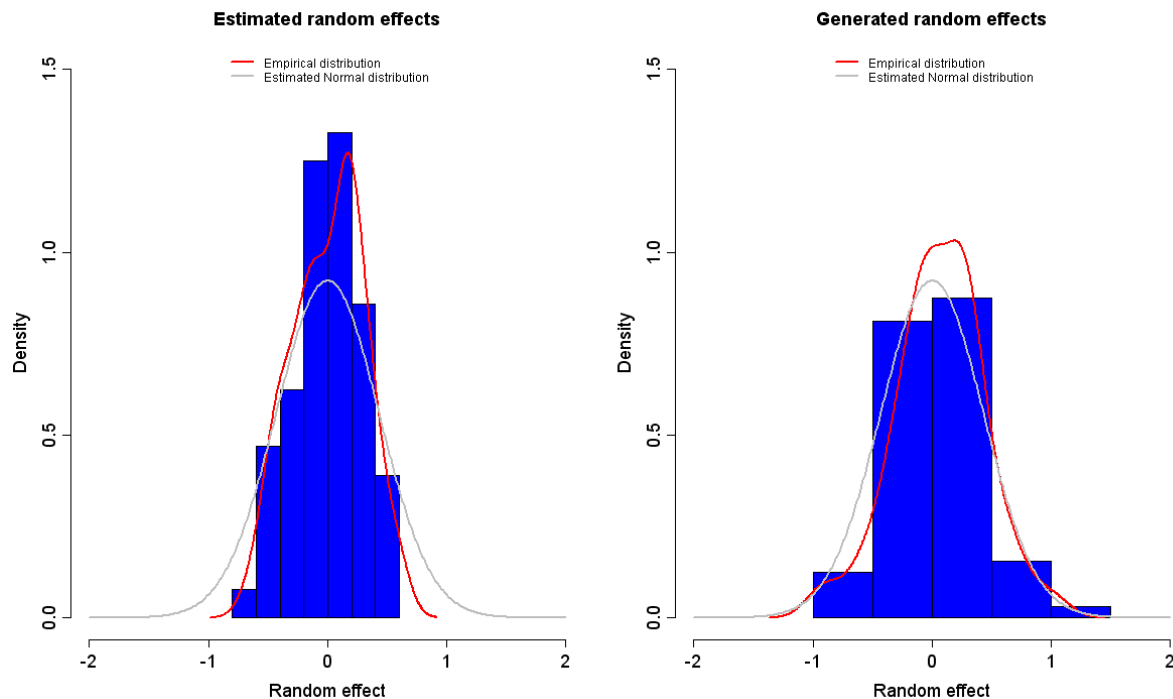


Figure 7: Distribution of estimated and generated random centre effects per centre

Simulations

The following steps are repeated $B = 1000$ times:

1. The centre choice for each patient i is simulated from a multinomial distribution using the estimated probabilities \widehat{PS} :

$$X_i \sim \text{Multinomial}(\widehat{PS}_{i1}, \widehat{PS}_{i2}, \dots, \widehat{PS}_{im}).$$

2. The outcome indicator (APR and Hartmann's procedure, yes (1) or no (0)) for each patient i in simulated centre c is simulated from a Bernoulli distribution:

$$Y_{ic} \sim \text{Bernoulli}\left(\text{expit}(\hat{\alpha} + \mathbf{L}_{ic}'\hat{\beta} + \psi_c^{\text{true}})\right),$$

with $\hat{\alpha}$ and $\hat{\beta}$ estimated in the true outcome model based on the original data.

The simulated 'success' rates were checked and are close the 'success' rate in the original Procure dataset.

3. The following statistical analyses are performed on each simulated dataset and per centre the mentioned outcome measures are stored for each dataset:

- O/E method (standardised event rate).
- Fixed-effects logistic regression model with Firth's bias correction likelihood penalty (odds ratio, centre-specific probability of success, population averaged probability of success and centre-specific standardised odds ratio).
- Hierarchical logistic regression model (odds ratio, centre-specific probability of success, population averaged probability of success and centre-specific standardised odds ratio).
- Propensity score method: regression doubly robust estimator (population averaged probability of success).

A summary of these outcome measures is graphically presented in Figures 16 - 25 in Appendix A.

Clearly, the empirical confidence intervals for the odds ratio's corresponding to the hierarchical logistic regression method are much smaller than those for the (Firth corrected) fixed-effects logistic regression model (Figure 8).

Tables 5 and 6 provide the statistical summary measures for the coverage of the empirical 95% confidence intervals (8) and the RMSE (9) per centre over all simulated datasets.

Method	Outcome measure	Min	25%-tile	Median	75%-tile	Max	Median width
O/E method	Standardised event rate (SER)	0	14	62	85	96	1.20
	Odds ratio	68	87	91	93	95	2.07
Fixed-effect logistic regression	Centre-specific probability of 'success'	72	85	88	91	95	0.37
	Population-averaged probability of 'success'	70	89	92	94	96	0.31
	(Centre-specific) standardised odds ratio	69	90	93	94	95	1.71
Hierarchical logistic regression	Odds ratio	0	67	89	94	96	0.63
	Centre-specific probability of 'success'	1	76	91	94	95	0.14
	Population-averaged probability of 'success'	0	74	91	94	96	0.12
	(Centre-specific) standardised odds ratio	0	66	89	94	96	0.61
Propensity score	Population-averaged probability of 'success'	72	92	94	95	95	0.39

Table 5: Statistical summary measures of the distribution over the centres of the coverage (%) of the empirical 95% confidence intervals over all simulated datasets for QCI 1232a (version 1), for all considered outcome measures

Method	Outcome measure	Min	25%-tile	Median	75%-tile	Max
O/E method	Standardised event rate (SER)	0.195	0.451	0.623	0.848	2.675
	Odds ratio	0.219	0.384	0.546	0.717	∞
Fixed-effect logistic regression	Centre-specific probability of 'success'	0.056	0.089	0.103	0.134	0.210
	Population-averaged probability of 'success'	0.049	0.076	0.089	0.110	0.189
	(Centre-specific) standardised odds ratio	0.194	0.335	0.446	0.603	∞
Hierarchical logistic regression	Odds ratio	0.116	0.159	0.236	0.336	1.445
	Centre-specific probability of 'success'	0.027	0.037	0.044	0.059	0.152
	Population-averaged probability of 'success'	0.026	0.035	0.041	0.055	0.146
	(Centre-specific) standardised odds ratio	0.114	0.157	0.233	0.330	1.368
Propensity score	Population-averaged probability of 'success'	0.038	0.076	0.104	0.133	0.293

Table 6: Statistical summary measures of the distribution over the centres of the RMSE over all simulated datasets for QCI 1232a (version 1), for all considered outcome measures

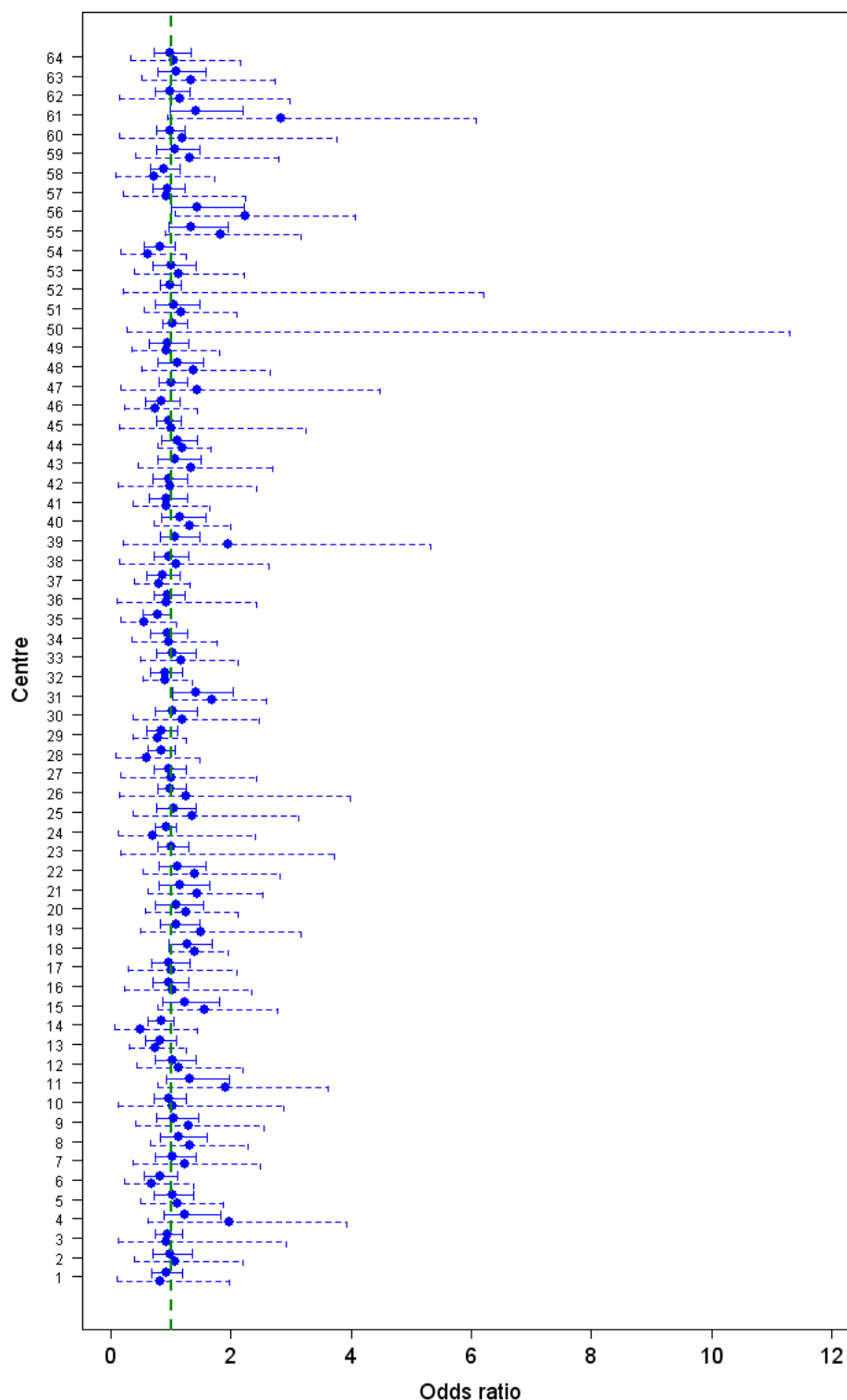


Figure 8: Forest plot to illustrate the shrinkage phenomenon for QCI 1232a (version 1). Blue dots represent the average of the simulated odds ratio's from the fixed-effects and hierarchical logistic regression model and empirical 95% confidence intervals for the fixed-effects logistic regression model are represented with a dotted line (–) and for the hierarchical logistic regression model with a full line (—).

4.2.2 Version 2: Incorporate misspecification of the random effects distribution

A breakpoint N_c is determined by estimating the random intercepts in a hierarchical logistic regression model fitted on the patients in small centres ($n_c < N_c$) and in a hierarchical logistic regression model fitted on the patients in large centres ($n_c \geq N_c$). An 'optimal' breakpoint for this exercise yields the largest discrepancy between the estimated intercepts in both models, based on this criterion it was decided to use the breakpoint $N_c = 31$.

Preparatory steps

- (a) A standard logistic regression model with only the patient-specific characteristics age, gender and cStage as covariates is fitted on the complete, original dataset and the predicted outcome (on the logit-scale)

$$\widehat{\text{logit}(Y)}_i = \hat{\alpha} + \mathbf{L}'_{ic}\beta$$

is to be used as offset further on.

- (b) A hierarchical logistic regression model with a random centre-effect and the offset $\widehat{\text{logit}(Y)}$ is fitted on the subset of ($n_1 = 549$) patients treated in small centres ($c : n_c < N_c$). From this model the estimated random intercept $\hat{\alpha}_1$ and estimated variance of the random effects $\hat{\sigma}_{\psi,1}^2$ will be used further on.
- (c) A hierarchical logistic regression model with a random centre-effect and the offset $\widehat{\text{logit}(Y)}$ is fitted on the subset of ($n_2 = 1806$) patients treated in large centres ($c : n_c \geq N_c$). From this model the estimated random intercept $\hat{\alpha}_2$ and variance of the random effects $\hat{\sigma}_{\psi,2}^2$ will be used further on.

The estimated intercepts $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are constrained⁴ to $\hat{\alpha}_1^*$ and $\hat{\alpha}_2^*$ in order to ensure that the average of the random centre effects to be generated will be 0.

Using the estimated $\hat{\sigma}_{\psi,1}^2$ and $\hat{\sigma}_{\psi,2}^2$ and the constrained $\hat{\alpha}_1^*$ and $\hat{\alpha}_2^*$ the 'true' centre effects are generated as follows

$$\psi_c^{\text{true}} \sim \begin{cases} \text{Normal}(\hat{\alpha}_1^*, \hat{\sigma}_{\psi,1}^2) & n_c < N_c \\ \text{Normal}(\hat{\alpha}_2^*, \hat{\sigma}_{\psi,2}^2) & n_c \geq N_c \end{cases}$$

The distribution of the BLUPs from the 'true' outcome model and of the misspecified 'true' centre effects (ψ_c^{true}) are shown in Figure 9.

Simulations

The following steps are repeated $B = 1000$ times:

1. The centre choice for each patient i is simulated from a multinomial distribution using the estimated probabilities \widehat{PS} :

$$X_i \sim \text{Multinomial}(\widehat{PS}_{i1}, \widehat{PS}_{i2}, \dots, \widehat{PS}_{im}).$$

⁴Criterion: $P(n_c < N_c)\hat{\alpha}_1^* + P(n_c \geq N_c)\hat{\alpha}_2^* = 0$, keeping the absolute distance between $\hat{\alpha}_1$ and $\hat{\alpha}_2$ fixed.

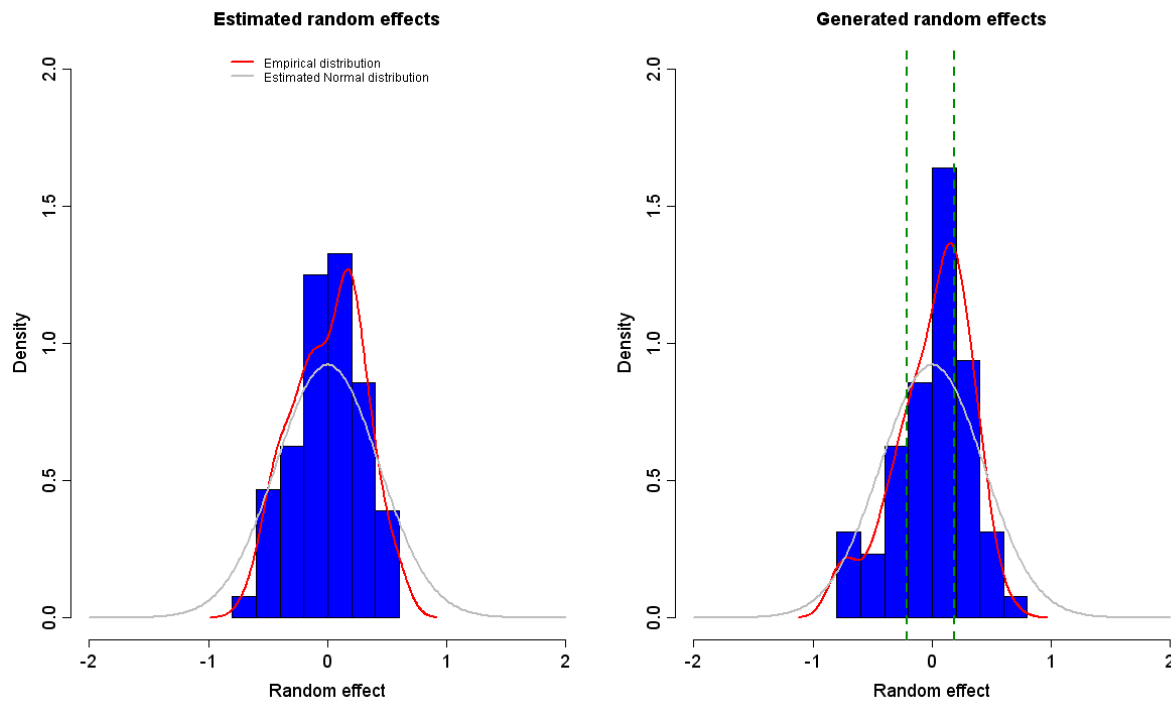


Figure 9: Distribution of estimated and generated random centre effects per centre, after constraining the average to be 0.

2. The outcome indicator (APR and Hartmann's procedure, yes or no) for each patient i in centre c is simulated from a Bernoulli distribution:

$$Y_{ic} \sim \text{Bernoulli} \left(\text{expit}(\widehat{\text{logit}}(Y)_i + \psi_c^{\text{true}}) \right).$$

The simulated 'success' rates were checked and are close to the 'success' rate in the original Procure dataset.

3. The following statistical analyses are performed on each simulated dataset and per centre the mentioned outcome measures are stored for each dataset:
 - O/E method (standardised event rate).
 - Fixed-effects logistic regression model with Firth's bias correction likelihood penalty (odds ratio, centre-specific probability of success, population averaged probability of success and centre-specific standardised odds ratio).
 - Hierarchical logistic regression model (odds ratio, centre-specific probability of success, population averaged probability of success and centre-specific standardised odds ratio).
 - Propensity score method: regression doubly robust estimator (population averaged probability of success).

A summary of these outcome measures is graphically presented in Figures 26 - 35 in Appendix B.

Clearly, the empirical confidence intervals for the odds ratio's corresponding to the hierarchical logistic regression method are much smaller than those for the (Firth corrected) fixed-effects logistic regression model.

Tables 7 and 8 provide the statistical summary measures for the coverage of the empirical 95% confidence intervals (8) and the RMSE (9) per centre over all simulated datasets.

Method	Outcome measure	Min	25%-tile	Median	75%-tile	Max	Median width
O/E method	Standardised event rate (SER)	0	50	80	93	96	1.31
	Odds ratio	71	86	89	91	95	1.86
Fixed-effect logistic regression	Centre-specific probability of 'success'	73	83	87	91	94	0.38
	Population-averaged probability of 'success'	70	87	91	93	96	0.31
	(Centre-specific) standardised odds ratio	74	90	92	93	95	1.60
	Odds ratio	32	84	92	95	96	0.71
Hierarchical logistic regression	Centre-specific probability of 'success'	43	86	92	94	97	0.14
	Population-averaged probability of 'success'	40	84	92	94	95	0.13
	(Centre-specific) standardised odds ratio	31	84	92	95	96	0.69
	Odds ratio	73	92	94	95	96	0.38

Table 7: Statistical summary measures of the distribution over the centres of the coverage (%) of the empirical 95% confidence intervals over all simulated datasets for QCI 1232a (version 2), for all considered outcome measures

Method	Outcome measure	Min	25%-tile	Median	75%-tile	Max
O/E method	Standardised event rate (SER)	0.123	0.414	0.587	0.796	1.853
	Odds ratio	0.142	0.349	0.505	0.806	∞
Fixed-effect logistic regression	Centre-specific probability of 'success'	0.059	0.094	0.112	0.136	0.212
	Population-averaged probability of 'success'	0.055	0.084	0.095	0.112	0.186
	(Centre-specific) standardised odds ratio	0.149	0.303	0.407	0.630	∞
	Odds ratio	0.120	0.196	0.222	0.261	0.458
Hierarchical logistic regression	Centre-specific probability of 'success'	0.023	0.038	0.043	0.051	0.083
	Population-averaged probability of 'success'	0.020	0.035	0.040	0.047	0.077
	(Centre-specific) standardised odds ratio	0.116	0.193	0.217	0.255	0.439
	Odds ratio	0.035	0.074	0.102	0.137	0.280

Table 8: Statistical summary measures of the distribution over the centres of the RMSE over all simulated datasets for QCI 1232a (version 2), for all considered outcome measures

4.3 Simulation results for the survival outcome: Overall 5-year survival by stage (QCI 1111)

This QCI is defined in Vlayen et al. (2008) as:

- Numerator: all RC patients that survived after 5 years, by stage.
- Denominator: all RC patients.
- Exclusion criteria:
 - patients treated abroad,
 - patients without a social security number,
 - patients without a Belgian postal code,
 - patients without a known incidence date or with an incidence date after December 31st, 2008.⁵

There are $n = 2294$ patients with full information on the patient-specific confounders and not meeting the exclusion criteria. After grouping small centres (with less than 5 patients meeting these criteria) into one overlapping centre, $m = 64$ centres are available for the simulation exercise (note that centre $c = 64$ is the overlapping centre). The propensity scores \widehat{PS} are estimated from a multinomial regression model with as covariates the main effects for age, gender and cStage (Figure 10). For some centres (e.g. 19, 41) there appears to be large variation between estimated propensities

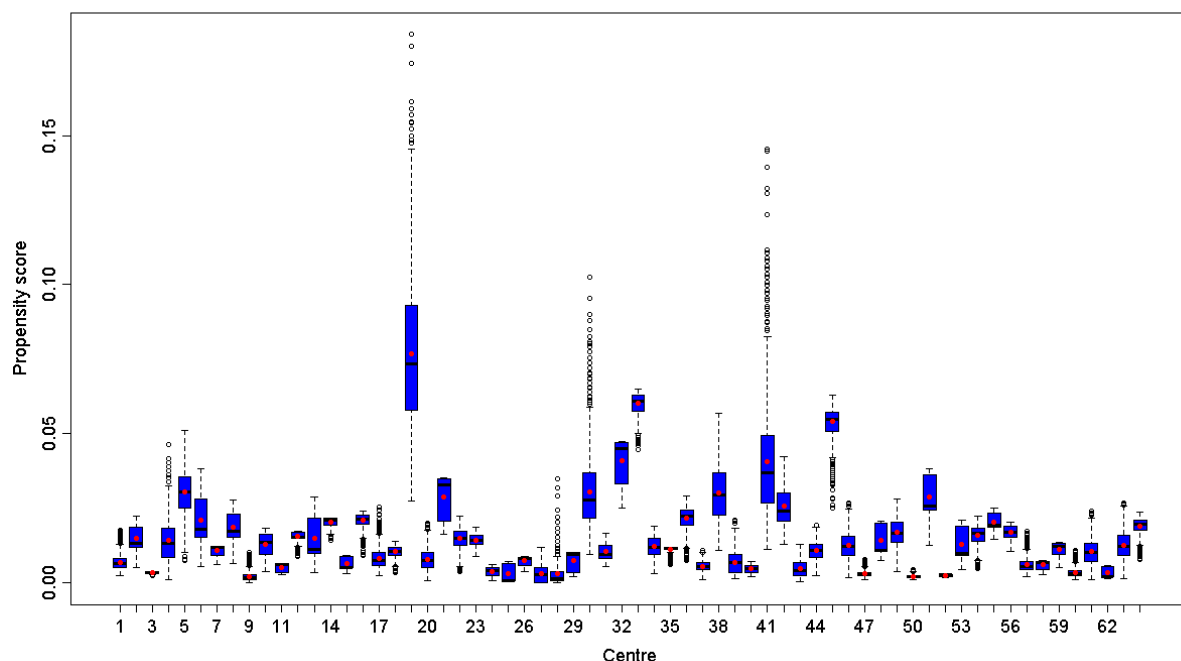


Figure 10: Distribution of estimated propensity scores per centre.

while for other centres (e.g. 3, 52) almost all patients have 0 probability of being treated there.

⁵The Procure database was linked with a version of the Cross Reference database of December 31st, 2008. In the Cross Reference database there still is a delay of about 6 months before guarantee of the complete information.

Overall, 12.6% of all patients did not survive December 31st, 2008, distribution of the outcome over the cStages is given in Table 9.

	$\Delta = 0$	$\Delta = 1$	% observed events
I	275	22	7.4%
II	374	47	11.2%
III	1071	100	8.5%
IV	207	110	34.7%
X	78	10	11.4%

Table 9: Distribution of outcome per cStage

The number of patients, person years and events per centre have been checked and no anomalies were shown [not shown to protect confidentiality].

The 'true' outcome model is the Cox proportional hazards frailty model, stratified by cStage

$$\lambda_i(t) = \lambda_{0,s_i}(t) \exp(\mathbf{L}'_i \boldsymbol{\beta} + \psi_c),$$

with $\exp(\Psi) \sim \text{Gamma}(1/\sigma_\psi^2, 1/\sigma_\psi^2)$, such that $E[\exp(\Psi)] = 1$ and $\text{Var}[\exp(\Psi)] = \sigma_\psi^2$.

Call:

```
coxph(formula = Surv(T, Delta) ~ L + strata(S) + frailty(X.orig))
```

```
n= 2294
      coef      se(coef) se2    Chisq DF   p
Lgender    -0.0764 0.1241  0.1229  0.38  1.0 0.5400
Lage         0.6110 0.0709  0.0704 74.36  1.0 0.0000
frailty(X.orig)                43.82 22.9 0.0054
```

```
      exp(coef) exp(-coef) lower .95 upper .95
Lgender    0.926      1.079    0.726    1.18
Lage        1.842      0.543    1.603    2.12
```

Iterations: 6 outer, 24 Newton-Raphson

Variance of random effect= 0.156 I-likelihood = -1589

Degrees of freedom for terms= 2.0 22.9

Rsquare= 0.066 (max possible= 0.761)

Likelihood ratio test= 157 on 24.9 df, p=0

Wald test = 74.4 on 24.9 df, p=7.85e-07

The baseline⁶ survival curves per stage are presented in Figure 11. From this we learn that we can reasonably estimate up to 2-year survival for all stages on the current data. This graph reflects what is well known and what we have seen in Table 9, that mostly patients with cStage = IV do not survive long.

Censoring appears to be strongly correlated with centre choice, this likely follows from the fact that not all centres started participating in Procure at the same time [output not shown].

⁶Baseline: a male patient of average age.

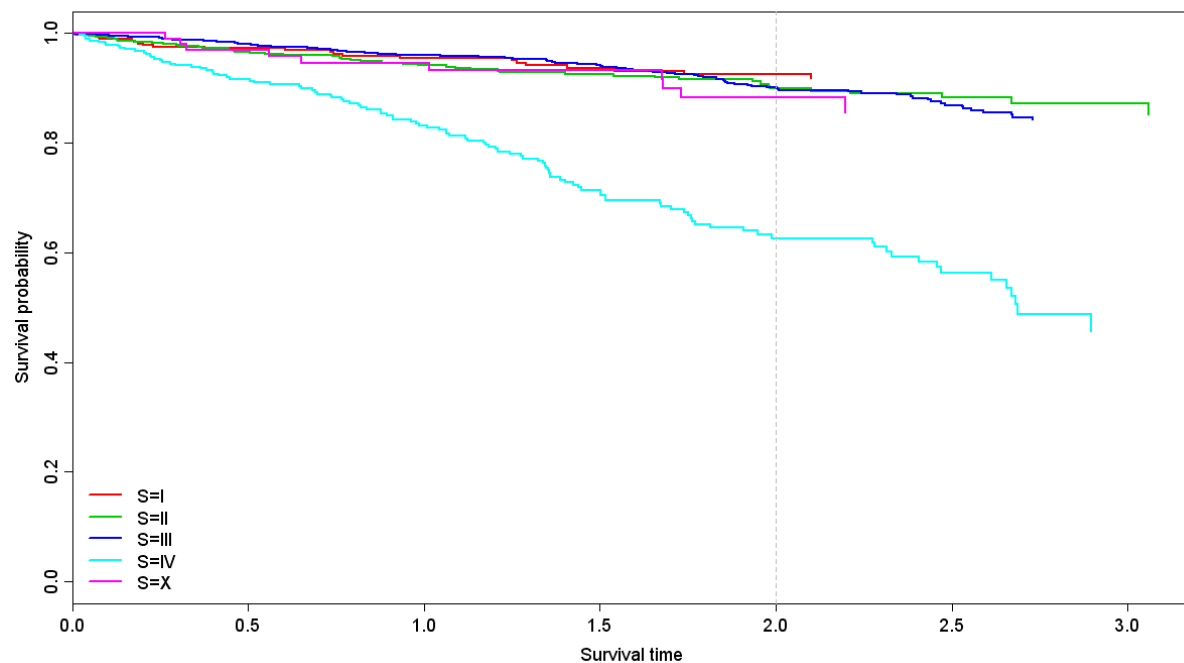


Figure 11: Baseline survival curves per stage, with baseline: male patients of average age.

Direct comparisons in performance between the different methods can already be made on the original data, namely by comparing the distribution of the estimated outcomes measures (the centre-specific and population averaged probability of 2-year survival and the centre-specific standardised odds ratio) between the different methods applied (Firth corrected Cox proportional hazards regression, frailty proportional hazards regression and the propensity score method), see Figures 12 and 13.

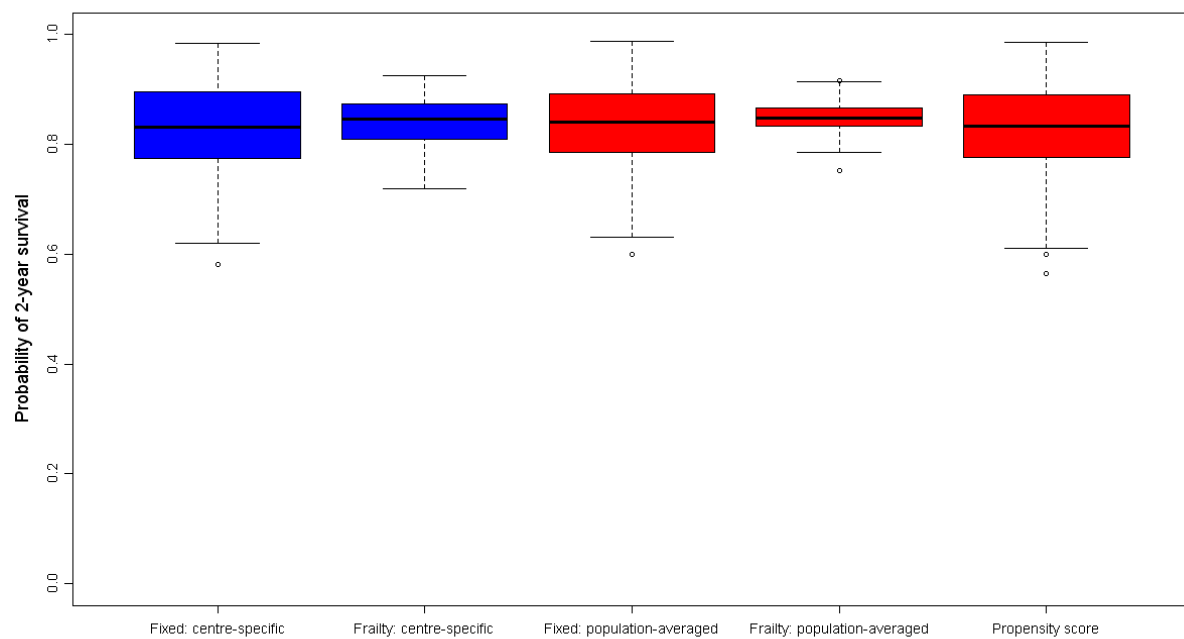


Figure 12: Comparison of centre-specific and population averaged probabilities of 2-year survival for QCI 1111 for the three applied statistical methods. The probabilities in this graph are estimated from the original Procure dataset.

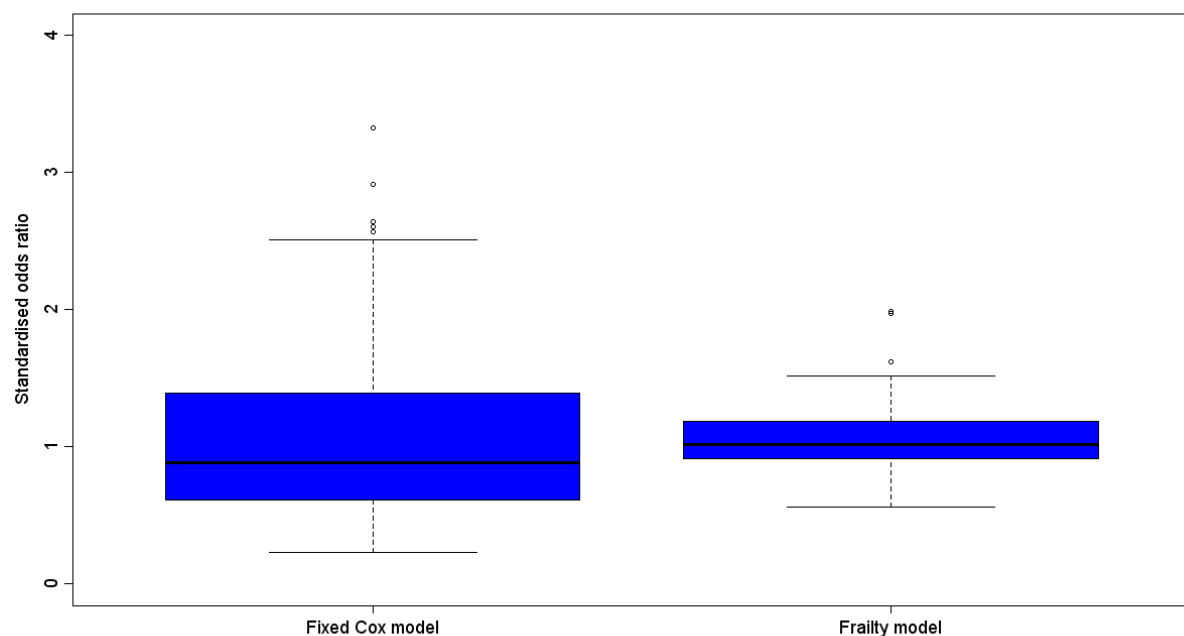


Figure 13: Comparison of the (centre-specific) standardised odds ratio of 2-year survival for QCI 1111 between the (Firth corrected) fixed-effects logistic regression and hierarchical logistic regression methods. The odds ratio's in this graph are estimated from the original Procure dataset.

4.3.1 Version 1: Use correct random effects distribution

Using the estimated $\hat{\sigma}_{\psi}^2$, the 'true' centre effects are generated as:

$$\exp(\psi_c^{\text{true}}) \sim \text{Gamma}(1/\hat{\sigma}_{\psi}^2, 1/\hat{\sigma}_{\psi}^2).$$

The distribution of the frailty terms as estimated on the original data and generated once for the simulations (ψ_c^{true}) are shown in Figure 14.

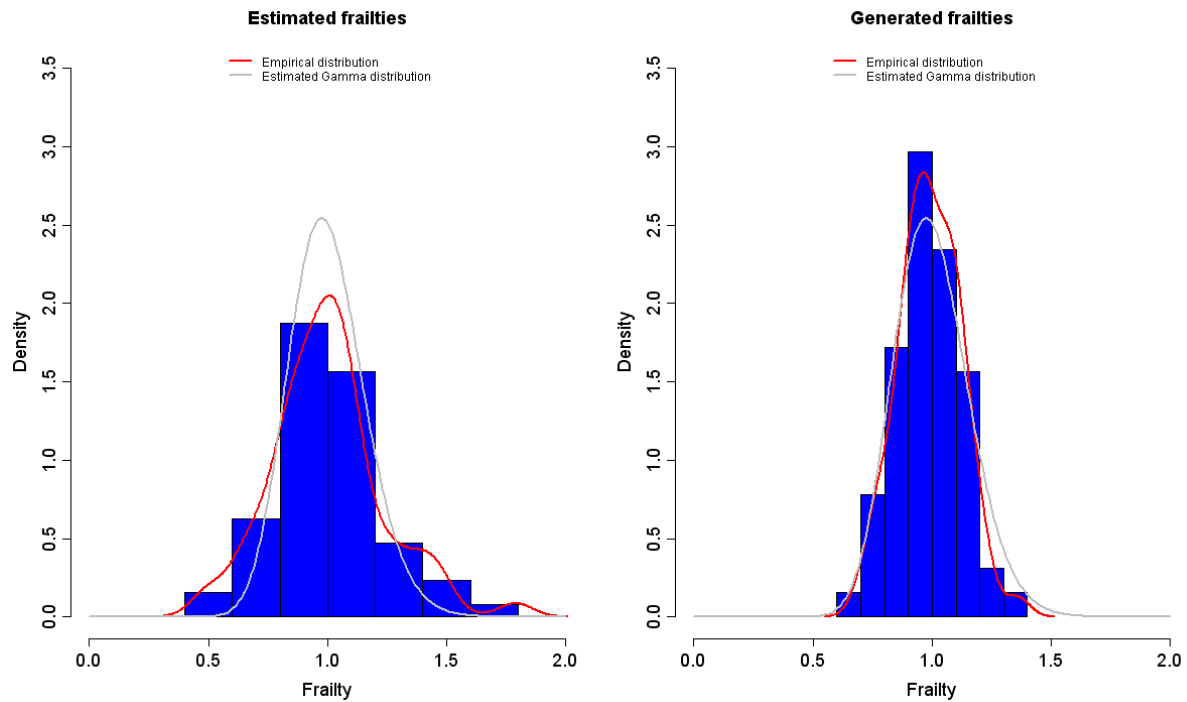


Figure 14: Distribution of the frailty terms as estimated on the original data (left panel) and generated once for the simulations (right panel)

Simulations

The following steps are repeated $B = 1000$ times:

1. Centre choice for each patient i is simulated from a multinomial distribution using the estimated probabilities \widehat{PS} :

$$X_i \sim \text{Multinomial}(\widehat{PS}_{i1}, \widehat{PS}_{i2}, \dots, \widehat{PS}_{im}).$$

2. The survival time for each patient i with cStage s_i in centre c is simulated from an exponential distribution:

$$D_i \sim \text{Exponential}(\lambda_{0,s_i} \exp(\mathbf{L}'_{ic} \boldsymbol{\beta} + \psi_{c_i})),$$

where the baseline hazards $\lambda_{0,s}$ are assumed to be constant and estimated as slope of a straight line through the cumulative baseline hazards, they are the number of events per year per cStage for a male patient of average age.

The event indicator is determined based on the fixed censoring time ($C_i = 31/12/2008 - \text{incidence date}$) and the simulated survival time, and simulated survival times are censored:

$$\Delta = \begin{cases} 0 & D_i > C_i \\ 1 & D_i \leq C_i \end{cases},$$

$$T_i = \min(D_i, C_i).$$

The simulated event times are similarly distributed as the original event times, but the simulated event rates are slightly higher (about 0.6%) than the event rate in the original Procure dataset.

3. Important note

It was observed that in many of the simulated datasets, the frailty proportional hazards model in R (version 2.10.1) estimated $\hat{\sigma}_\psi^2$ to be 0, which is not correct under the applied simulation scheme. To avoid including such simulated dataset in our simulation exercise the criterion $\hat{\sigma}_\psi^2 \geq 0.1$ was imposed before a dataset was actually withheld for the simulations. This value was chosen to at least approximate the variability random centre effects seen in the original Procure database (where $\hat{\sigma}_\psi^2 = 0.156$).

4. The following statistical analyses are performed on each simulated dataset and per centre the following outcome measures are stored for each dataset:

- Cox proportional hazards model with Firth's bias correction likelihood penalty (centre-specific, population averaged 2-year survival probability and standardised odds ratio).
- The frailty proportional hazards model (centre-specific, population averaged 2-year survival probability and standardised odds ratio).
- The propensity score method (doubly robust estimator of population averaged 2-year survival probability).

A summary of these outcome measures is graphically presented in Figures 36 - 42 in Appendix C.

Clearly, the 95% empirical confidence intervals for the odds ratio's corresponding to the hierarchical logistic regression method are much smaller than those for the (Firth corrected) fixed-effects logistic regression model.

Tables 10 and 11 provide the statistical summary measures for the coverage of the empirical 95% confidence intervals (8) and the RMSE (9) per centre over all simulated datasets (Figure 15).

Method	Outcome measure	Min	25%-tile	Median	75%-tile	Max	Median width
Cox proportional hazards regression	Centre-specific probability of 2-year survival	65	82	91	94	96	0.33
	Population-averaged probability of 2-year survival	62	80	92	94	96	0.29
	(Centre-specific) standardised odds ratio	10	82	90	93	95	0.51
Frailty proportional hazards regression	Centre-specific probability of 2-year survival	66	88	91	94	96	0.15
	Population-averaged probability of 2-year survival	54	87	92	94	95	0.10
	(Centre-specific) standardised odds ratio	60	85	88	92	95	0.77
Propensity score	Population-averaged probability of 2-year survival	63	79	88	90	94	0.33

Table 10: Statistical summary measures of the distribution over the centres of the coverage (%) of the empirical 95% confidence intervals over all simulated datasets for QCI 1111 (version 1), for all considered outcome measures

Method	Outcome measure	Min	25%-tile	Median	75%-tile	Max
Cox proportional hazards regression	Centre-specific probability of 2-year survival	0.031	0.069	0.087	0.110	0.220
	Population-averaged probability of 2-year survival	0.035	0.065	0.079	0.103	0.247
	(Centre-specific) standardised odds ratio	0.086	0.131	0.177	0.244	0.415
Frailty proportional hazards regression	Centre-specific probability of 2-year survival	0.023	0.034	0.041	0.047	0.080
	Population-averaged probability of 2-year survival	0.015	0.026	0.028	0.030	0.037
	(Centre-specific) standardised odds ratio	0.098	0.191	0.218	0.241	0.310
Propensity score	Population-averaged probability of 2-year survival	0.040	0.074	0.092	0.119	0.260

Table 11: Statistical summary measures of the distribution over the centres of the RMSE over all simulated datasets for QCI 1111 (version 1), for all considered outcome measures

Note

In some of the simulate dataset the following problems are observed to occur:

- There are no follow-up times beyond 2 years in one (or more) specific stratum (strata), hence it is not possible to provide an (stratum-corrected) estimate of the probability of 2-year survival.
- The loglikelihood-optimising algorithm for the (Firth-corrected) Cox proportional hazards regression in SAS (version 9.2) does not converge and no parameter estimates were obtained, hence no estimate of the probability of 2-year survival can be computed.

These problems are not entirely unexpected given the limited information (in terms of follow-up) available in the current Procure database. We plan to re-do this simulation exercise for QCI 1111 when an updated version of the database (relevant for the statistical analysis to be performed) is available.

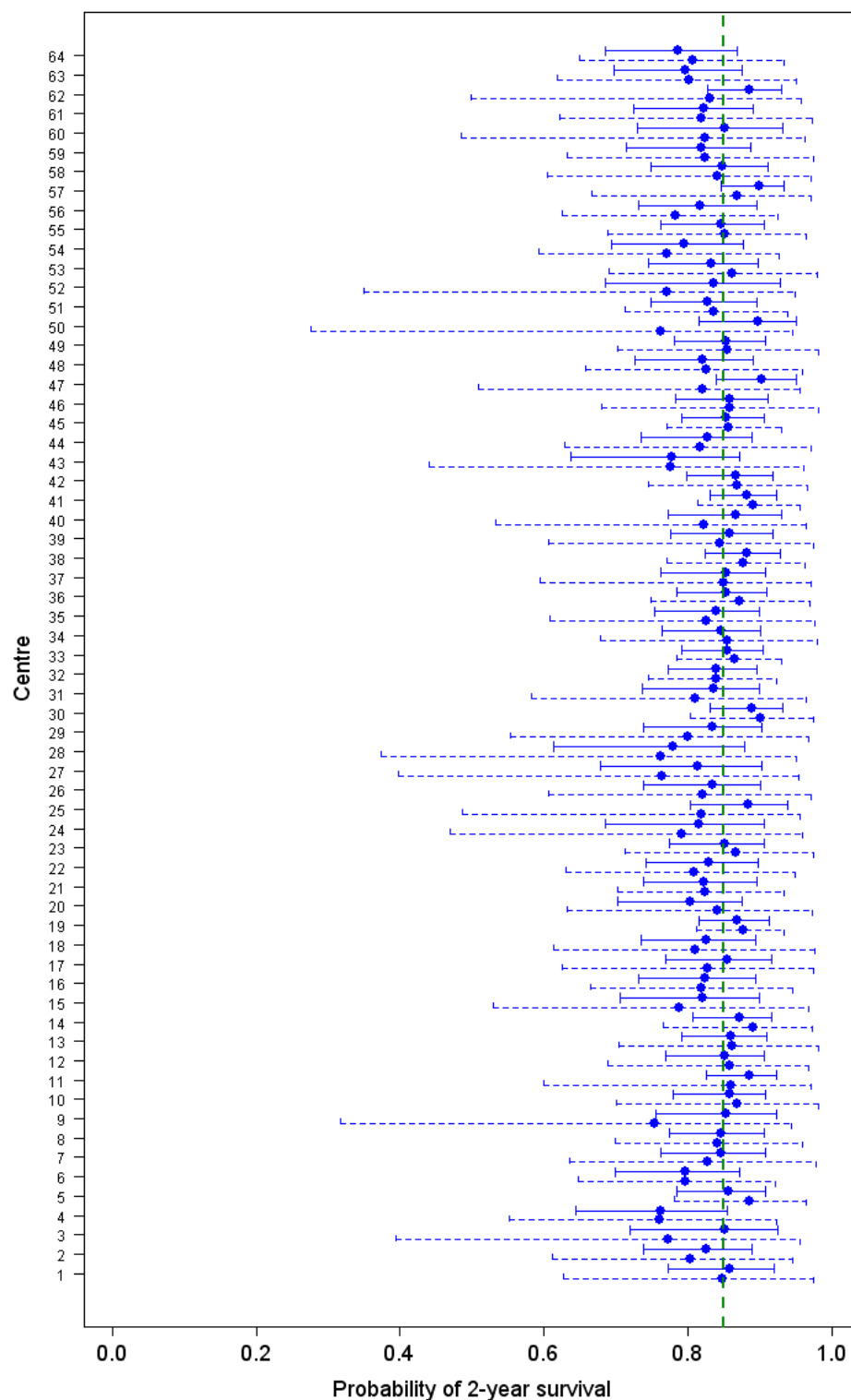


Figure 15: Forest plot to illustrate the shrinkage phenomenon for QCI 1111. Blue dots represent the average of the simulated centre-specific probabilities of 2-year survival from the Cox proportional hazards model and frailty Cox proportional hazards model and empirical 95% confidence intervals for the Cox proportional hazards model are represented with a dotted line (–) and for the frailty Cox proportional hazards model with a full line (–). The green vertical line represents the overall probability of 2-year survival in the original Procure database.

References

- AUSTIN, P., ALTER, D. and TU, J. (2003). The use of fixed- and random-effects models for classifying hospitals as mortality outliers: A monte carlo assessment. *Medical Decision Making*, **23** 526–539.
- COLE, S. and HERNAN, M. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology*, **168** 656–664.
- DELONG, E., PETERSON, E., DELONG, D., MUHLBAIER, L., HACKETT, S. and MARK, D. (1997). Comparing risk-adjustment methods for provider profiling. *Statistics in Medicine*, **16** 2645–2664.
- FARIS, P., GHALI, W. and BRANT, R. (2003). Bias in estimates of confidence intervals for health outcome report cards. *Journal of Clinical Epidemiology*, **56** 553–558.
- FEDELI, U., BROCCO, S., ALBA, N., ROSATO, R. and SPOLAORE, P. (2007). The choice between different statistical approaches to risk-adjustment influenced the identification of outliers. *Journal of Clinical Epidemiology*, **60** 858–862.
- FIRTH, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, **80** 27–38.
- GOETGELUK, S., VANSTEELANDT, S. and GOETGHEBEUR, E. (2008). Estimation of controlled direct effects. *Journal of the Royal Statistical Society - Series B*, **70** 1049–1066.
- HARRELL, F. E. J. (2001). *Regression Modeling Strategies*. Springer Series in Statistics.
- HEINZE, G. and SCHEMPER, M. (2001). A solution to the problem of Monotone Likelihood in Cox Regression. *Biometrics*, **57** 114–119.
- HEINZE, G. and SCHEMPER, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, **21** 2409–2419.
- ROBINS, J., SUED, M., LEI-GOMEZ, Q. and ROTNITZKY, A. (2008). Performance of double-robust estimators when inverse probability weights are highly variable. *Statistical Science*, **22** 544–559.
- TSIATIS, A., DAVIDIAN, M., ZHANG, M. and LU, X. (2008). Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. *Statistics in Medicine*, **27** 4658–4677.
- VANSTEELANDT, S., BEKAERT, M. and CLAESKENS, G. (2010). On model selection and model misspecification in causal inference. *Technical report*.
- VLAYEN, J., VERSTREKEN, M., MERTENS, C., VAN EYCKEN, E. and PENNINGCKX, F. (2008). Quality insurance for rectal cancer - phase 2: Development and testing of a set of quality indicators. *Good Clinical Practice (GCP. Brussels, Belgian Health Care Knowledge Centre (KCE)*.

A Evaluation of simulations for QCI 1232a, version 1

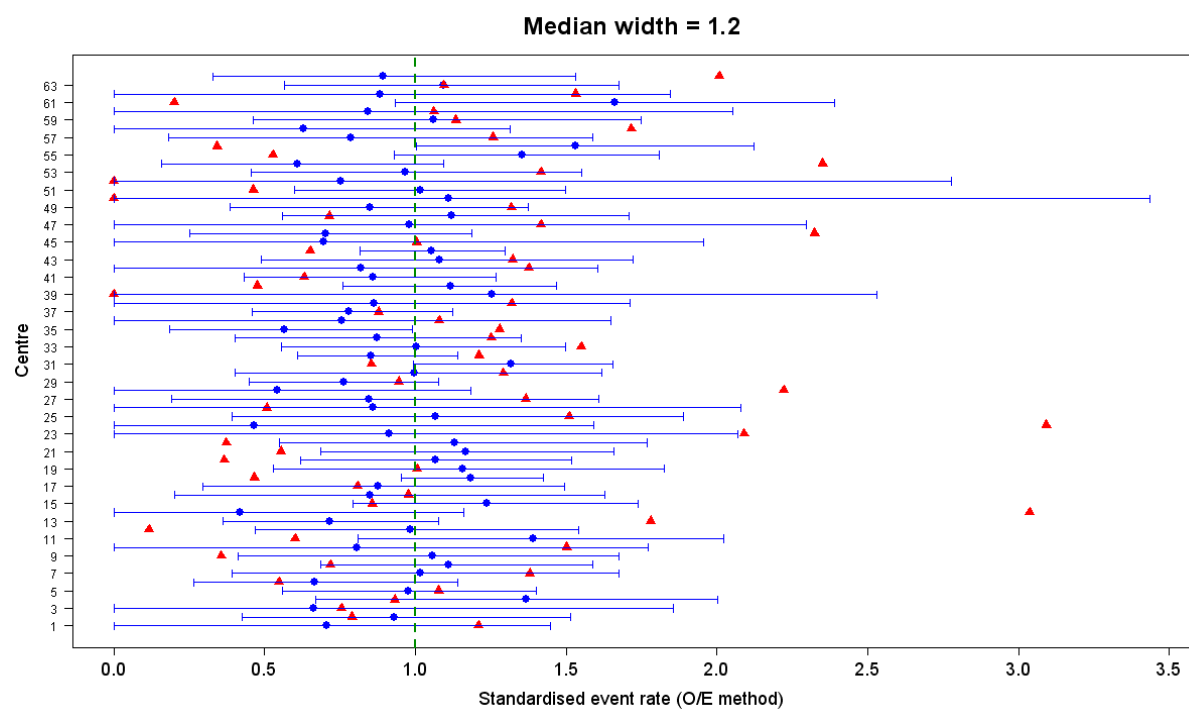


Figure 16: On QCI 1232a (version 1): standardised event rates for the **O/E method** per centre. The red triangles represent the 'true' standardised event rates, the blue bullets represent the average standardised event rates over 1000 simulations and the intervals are based on the empirical distribution of all simulated standardised event rates.

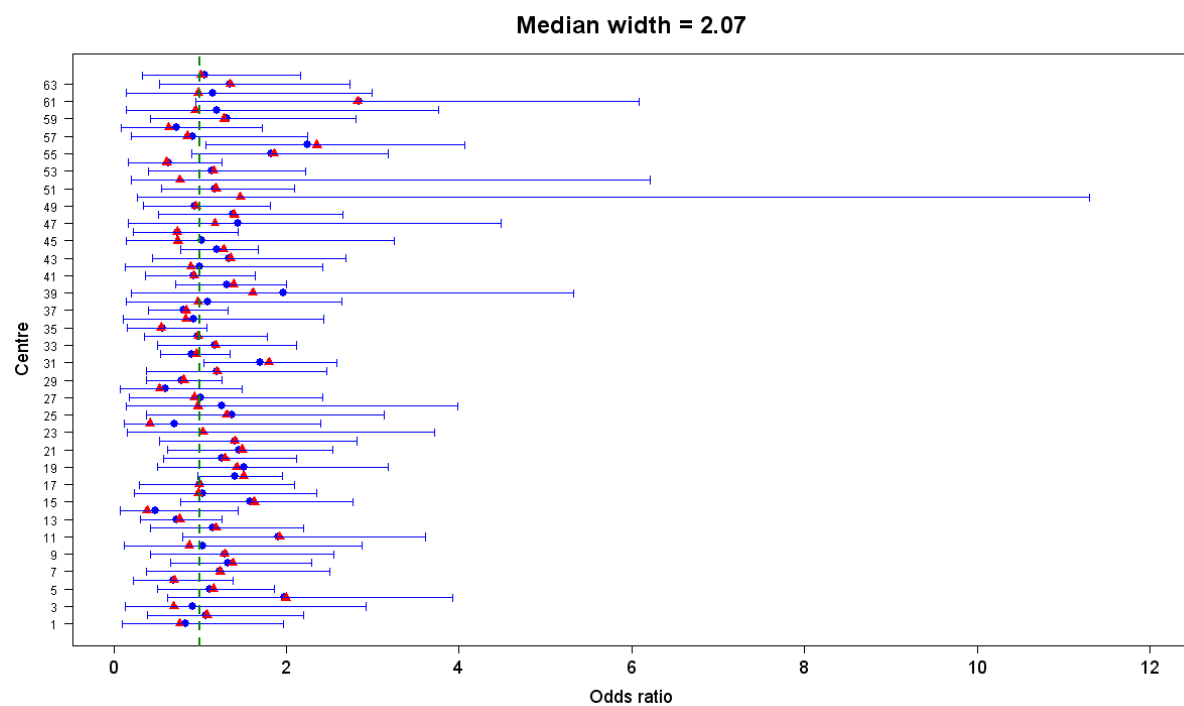


Figure 17: On QCI 1232a (version 1): odds ratio for the **(Firth corrected) fixed effects logistic regression** model, per centre. The red triangles represent the 'true' odds ratio, the blue bullets represent the average odds ratio's over 1000 simulations and the intervals are based on the empirical distribution of all odds ratio's per centre.

Note: for the centres without a blue bullet, the average odds ratio over all simulations was beyond the scope of this graph. Instead the median odds ratio over all simulations could be reported.

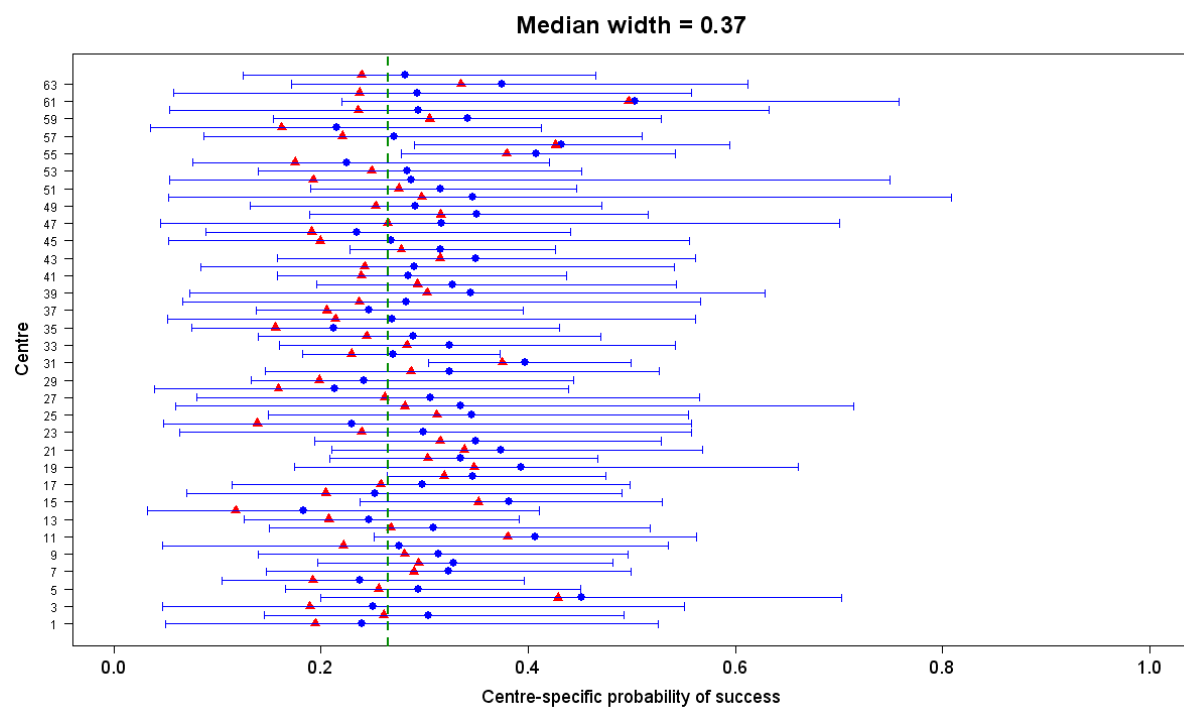


Figure 18: On QCI 1232a (version 1): centre-specific probability of success for the **(Firth corrected) fixed effects logistic regression** model, per centre. The red triangles represent the 'true' centre-specific probability of success, the blue bullets represent the average centre-specific probabilities of success over 1000 simulations and the intervals are based on the empirical distribution of all simulated centre-specific probabilities of success.

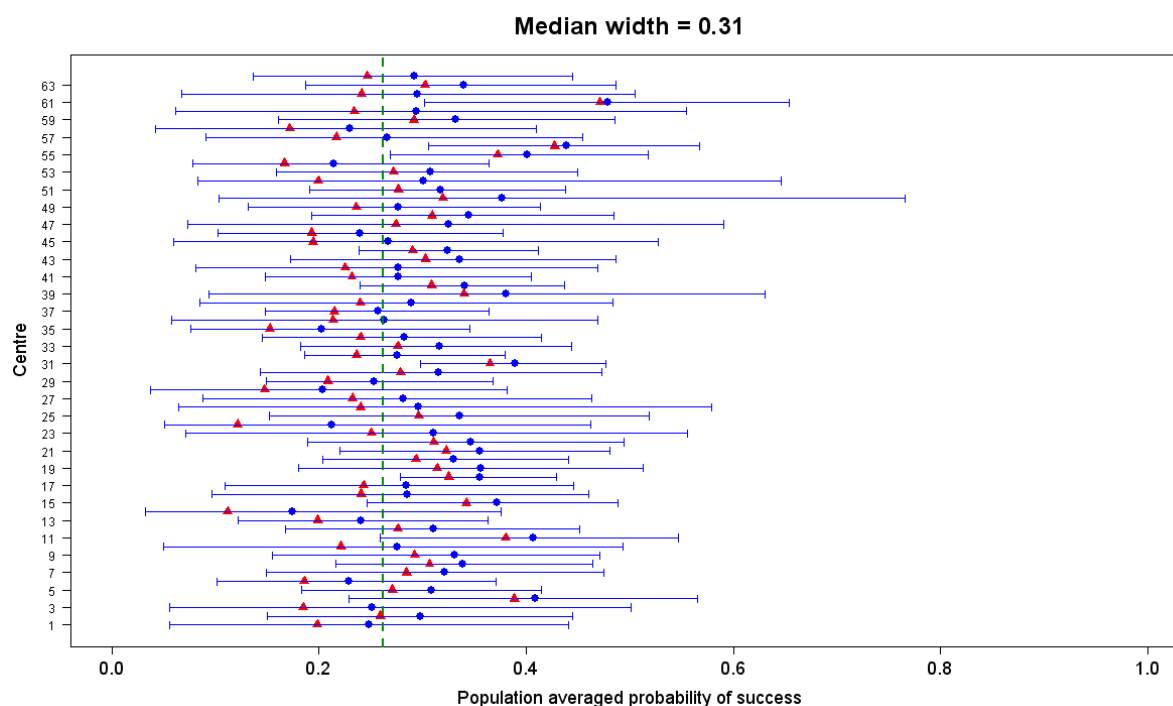


Figure 19: On QCI 1232a (version 1): population averaged probability of success for the **(Firth corrected) fixed effects logistic regression** model, per centre. The red triangles represent the 'true' population averaged probability of success, the blue bullets represent the average population averaged probabilities of success over 1000 simulations and the intervals are based on the empirical distribution of all simulated population averaged probabilities of success.

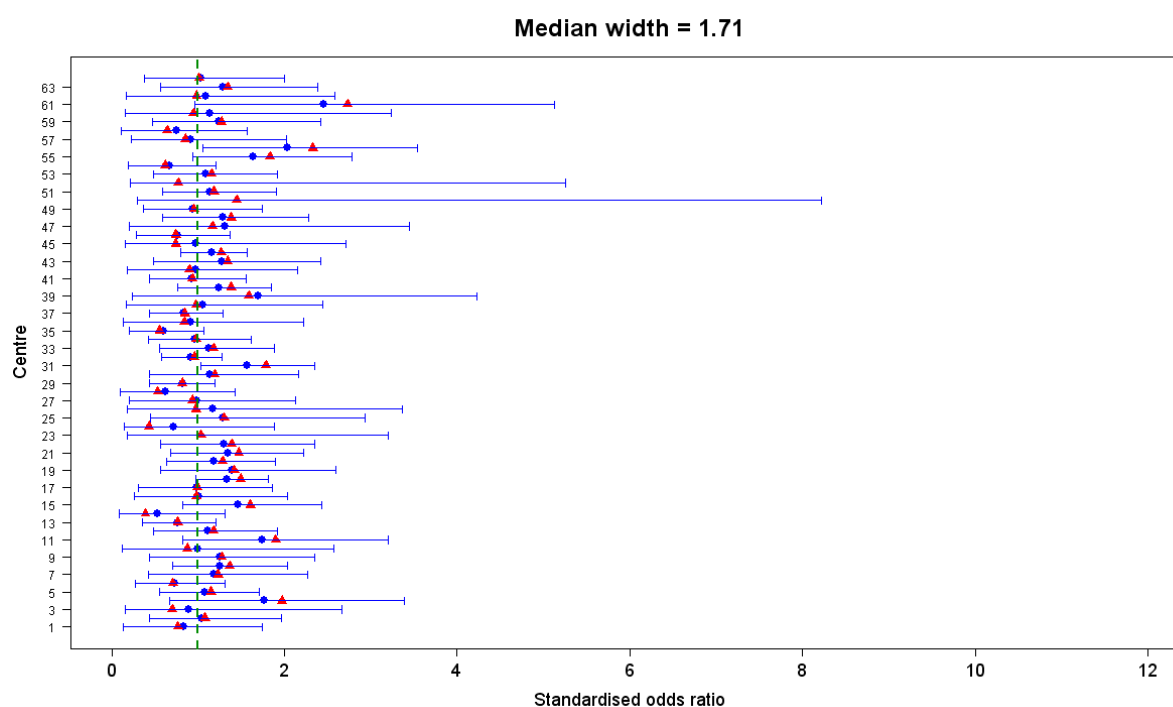


Figure 20: On QCI 1232a (version 1): (centre-specific) standardised odds ratio for the **(Firth corrected) fixed effects logistic regression** model, per centre. The red triangles represent the 'true' standardised odds ratio's, the blue bullets represent the average standardised odds ratio's over 1000 simulations and the intervals are based on the empirical distribution of all simulated odds ratio's.

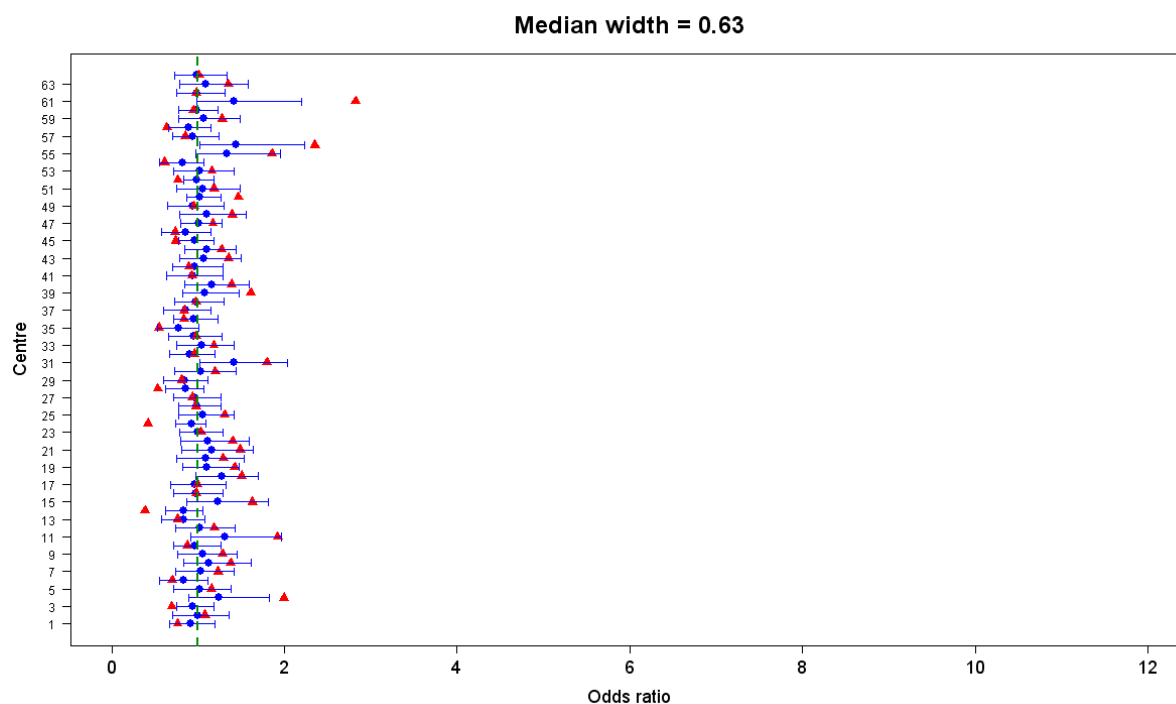


Figure 21: On QCI 1232a (version 1): odds ratio for the **hierarchical logistic regression** model, per centre. The red triangles represent the 'true' odds ratio's, the blue bullets represent the average odds ratio's over 1000 simulations and the intervals are based on the empirical distribution of all simulated odds ratio's.

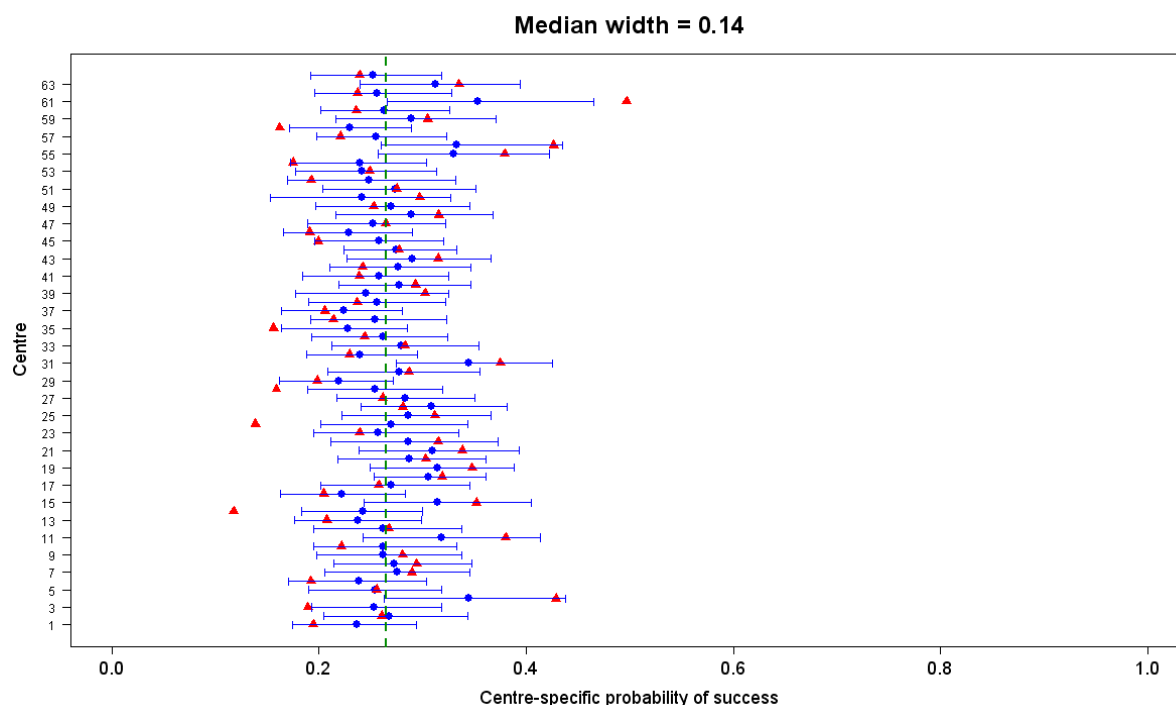


Figure 22: On QCI 1232a (version 1): centre-specific probability of success for the **hierarchical logistic regression** model, per centre. The red triangles represent the 'true' centre-specific probabilities of success, the blue bullets represent the average centre-specific probabilities of success over 1000 simulations and the intervals are based on the empirical distribution of all simulated centre-specific probabilities of success.

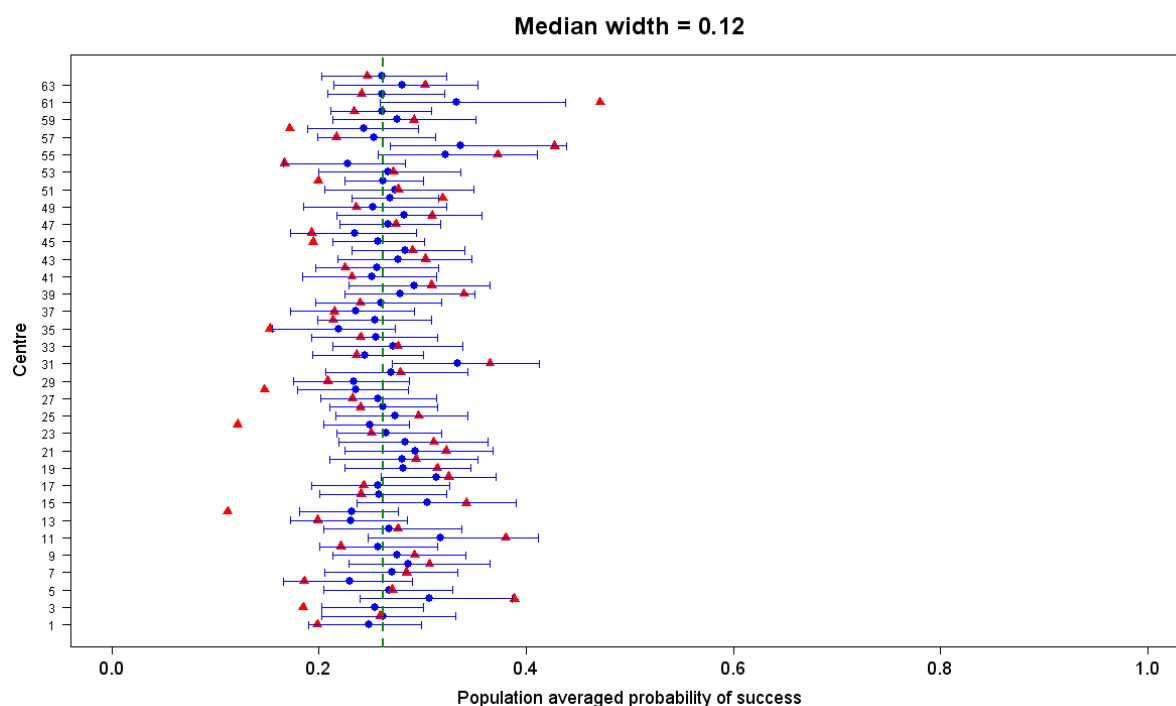


Figure 23: On QCI 1232a (version 1): population averaged probability of success for the **hierarchical logistic regression** model, per centre. The red triangles represent the 'true' population averaged probabilities of success, the blue bullets represent the average population averaged probabilities of success over 1000 simulations and the intervals are based on the empirical distribution of all simulated population averaged probabilities of success.

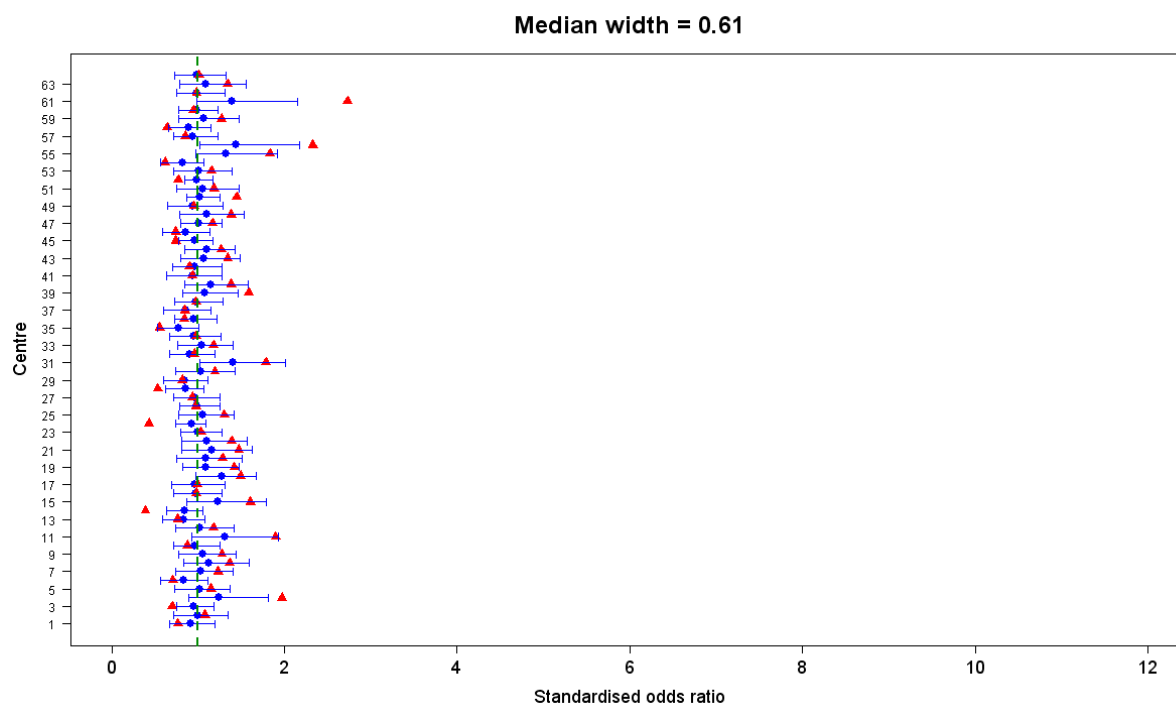


Figure 24: On QCI 1232a (version 1): (centre-specific) standardised odds ratio for the **hierarchical logistic regression** model, per centre. The red triangles represent the 'true' standardised odds ratio's, the blue bullets represent the average standardised odds ratio's over 1000 simulations and the intervals are based on the empirical distribution of all simulated odds ratio's.

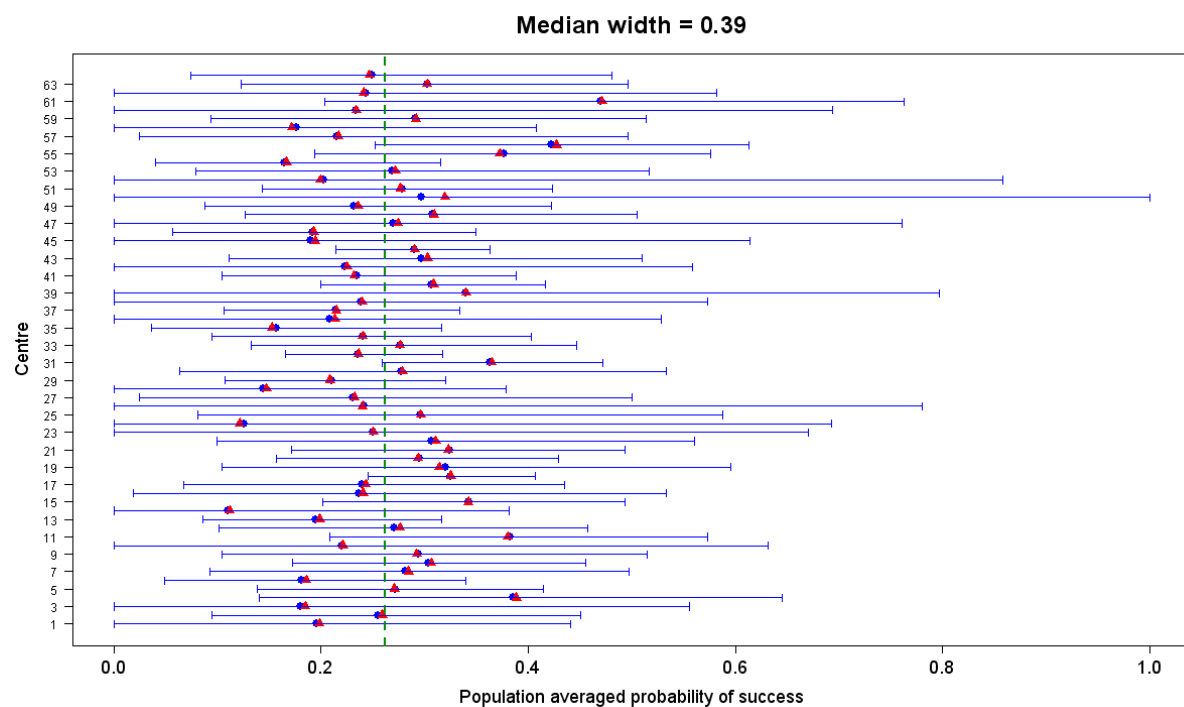


Figure 25: On QCI 1232a (version 1): population averaged probability of success for the **propensity score method**, per centre. The red triangles represent the 'true' population averaged probabilities of success, the blue bullets represent the average population averaged probabilities of success over 1000 simulations and the intervals are based on the empirical distribution of all simulated population averaged probabilities of success.

B Evaluation of simulations for QCI 1232a, version 2

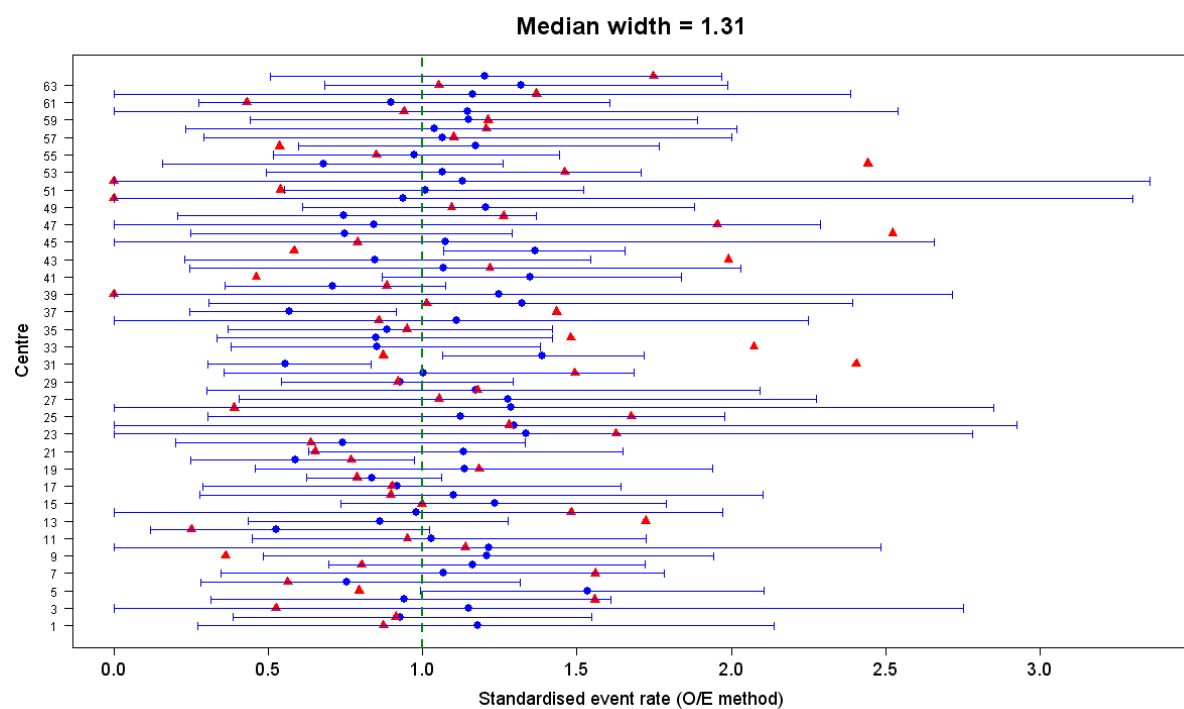


Figure 26: On QCI 1232a (version 2): standardised event rate for the **O/E method** per centre. The red triangles represent the 'true' standardised event rate, the blue bullets represent the average standardised event rate over 1000 simulations and the intervals are based on the empirical distribution of all simulated standardised event rates.

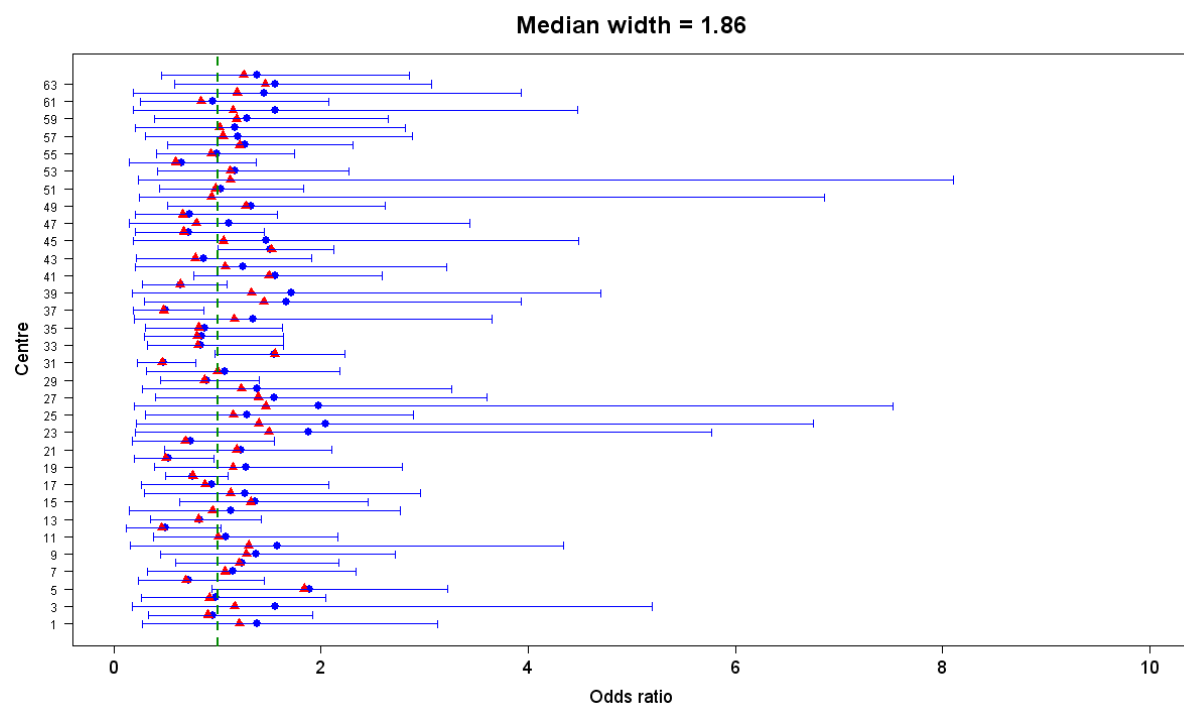


Figure 27: On QCI 1232a (version 2): odds ratio for the **(Firth corrected) fixed effects logistic regression** model, per centre. The red triangles represent the 'true' odds ratio, the blue bullets represent the average odds ratio's over 1000 simulations and the intervals are based on the empirical distribution of all odds ratio's per centre.

Note: for the centres without a blue bullet, the average odds ratio over all simulations was beyond the scope of this graph. Instead the median odds ratio over all simulations could be reported.

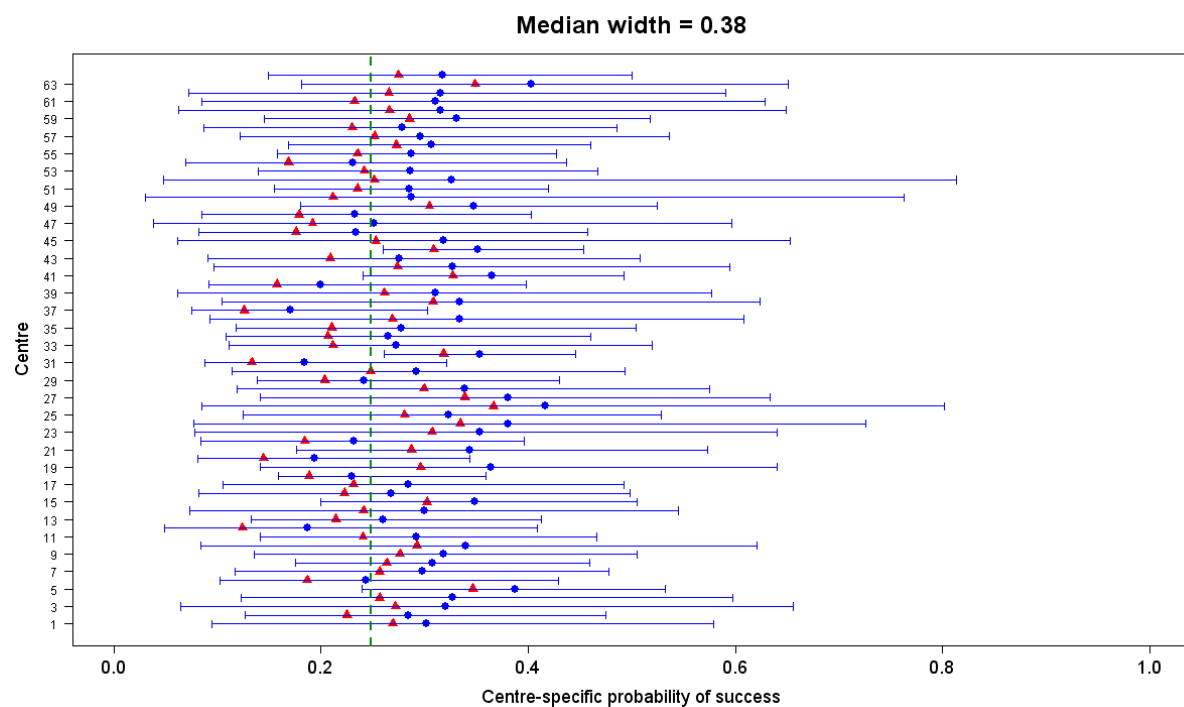


Figure 28: On QCI 1232a (version 2): centre-specific probability of success for the **(Firth corrected) fixed effects logistic regression** model, per centre. The red triangles represent the 'true' centre-specific probability of success, the blue bullets represent the average centre-specific probabilities of success over 1000 simulations and the intervals are based on the empirical distribution of all simulated centre-specific probabilities of success.

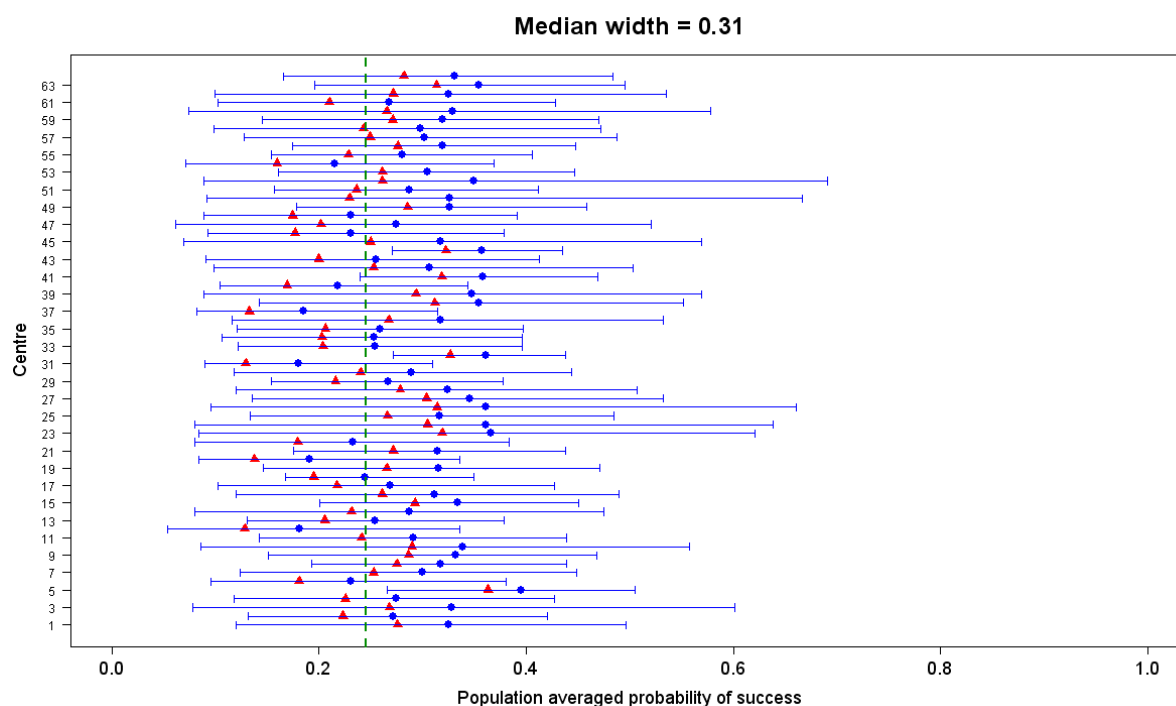


Figure 29: On QCI 1232a (version 2): population averaged probability of success for the **(Firth corrected) fixed effects logistic regression** model, per centre. The red triangles represent the 'true' population averaged probability of success, the blue bullets represent the average population averaged probabilities of success over 1000 simulations and the intervals are based on the empirical distribution of all simulated population averaged probabilities of success.

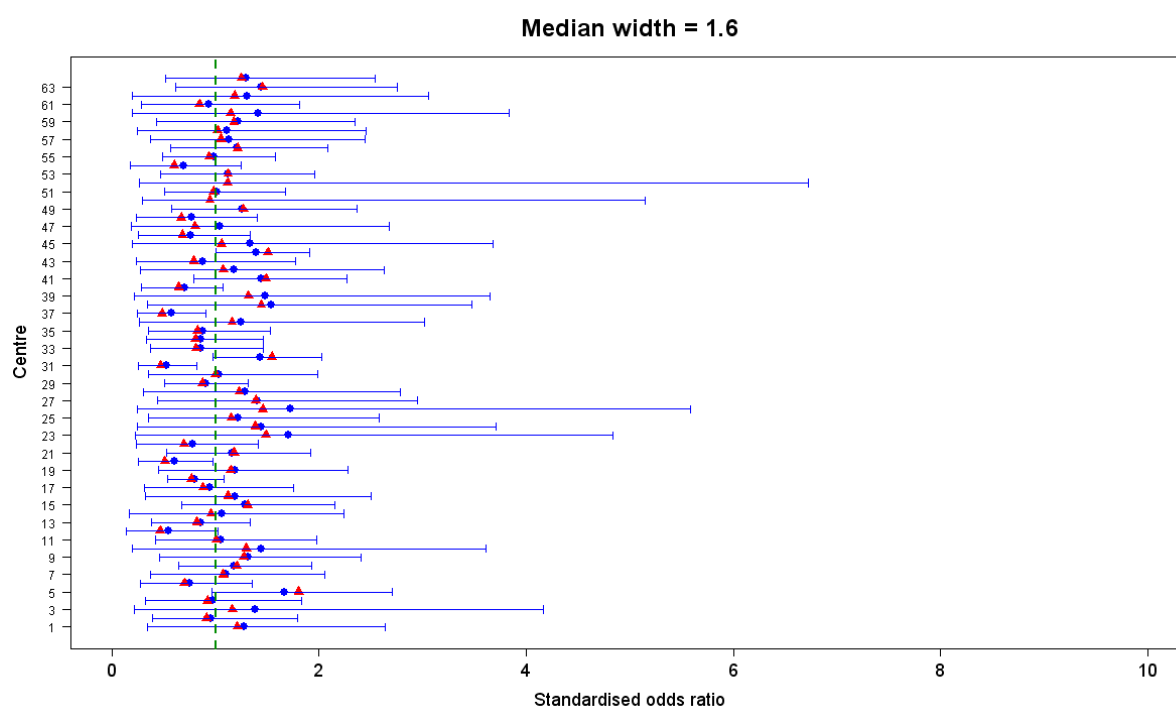


Figure 30: On QCI 1232a (version 2): (centre-specific) standardised odds ratio for the **(Firth corrected) fixed effects logistic regression** model, per centre. The red triangles represent the 'true' standardised odds ratio's, the blue bullets represent the average standardised odds ratio's over 1000 simulations and the intervals are based on the empirical distribution of all simulated odds ratio's.

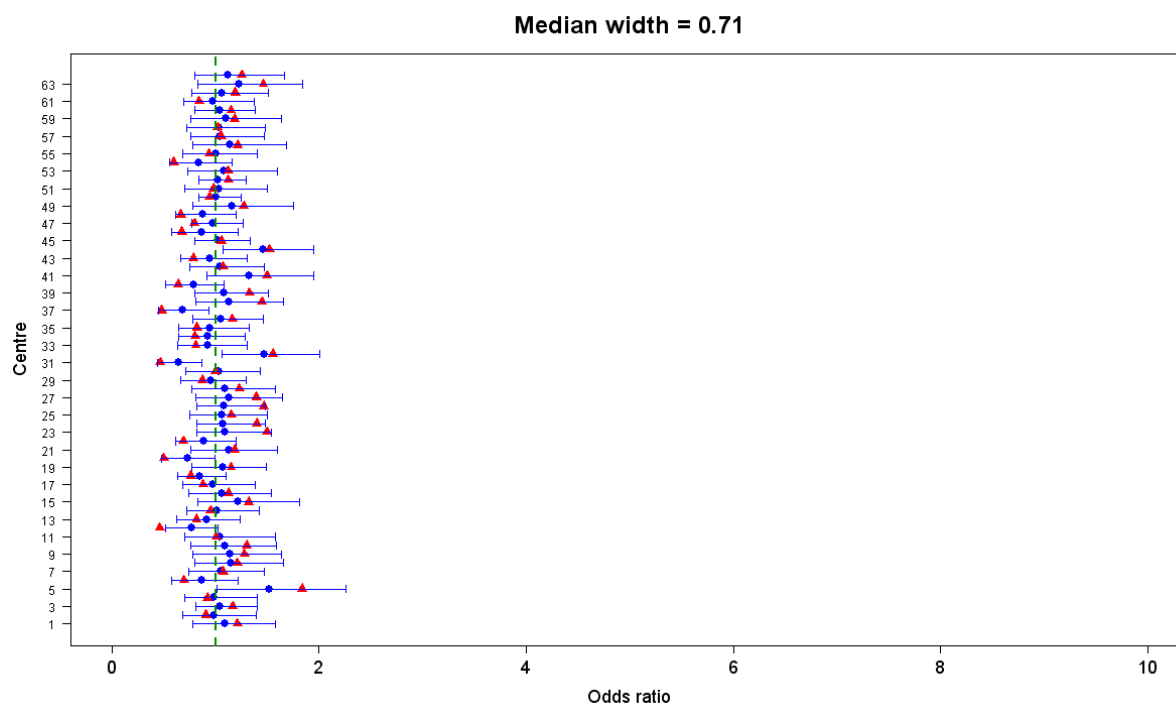


Figure 31: On QCI 1232a (version 2): odds ratio for the **hierarchical logistic regression** model, per centre. The red triangles represent the 'true' odds ratio's, the blue bullets represent the average odds ratio's over 1000 simulations and the intervals are based on the empirical distribution of all simulated odds ratio's.

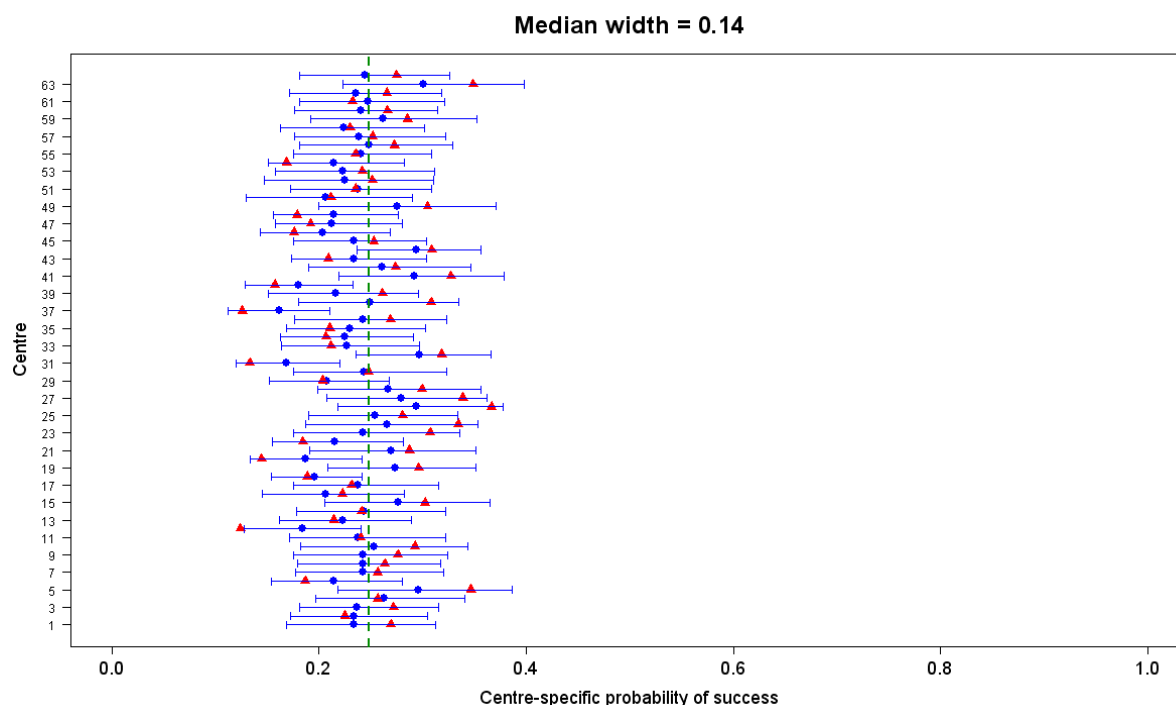


Figure 32: On QCI 1232a (version 2): centre-specific probability of success for the **hierarchical logistic regression** model, per centre. The red triangles represent the 'true' centre-specific probabilities of success, the blue bullets represent the average centre-specific probabilities of success over 1000 simulations and the intervals are based on the empirical distribution of all simulated centre-specific probabilities of success.

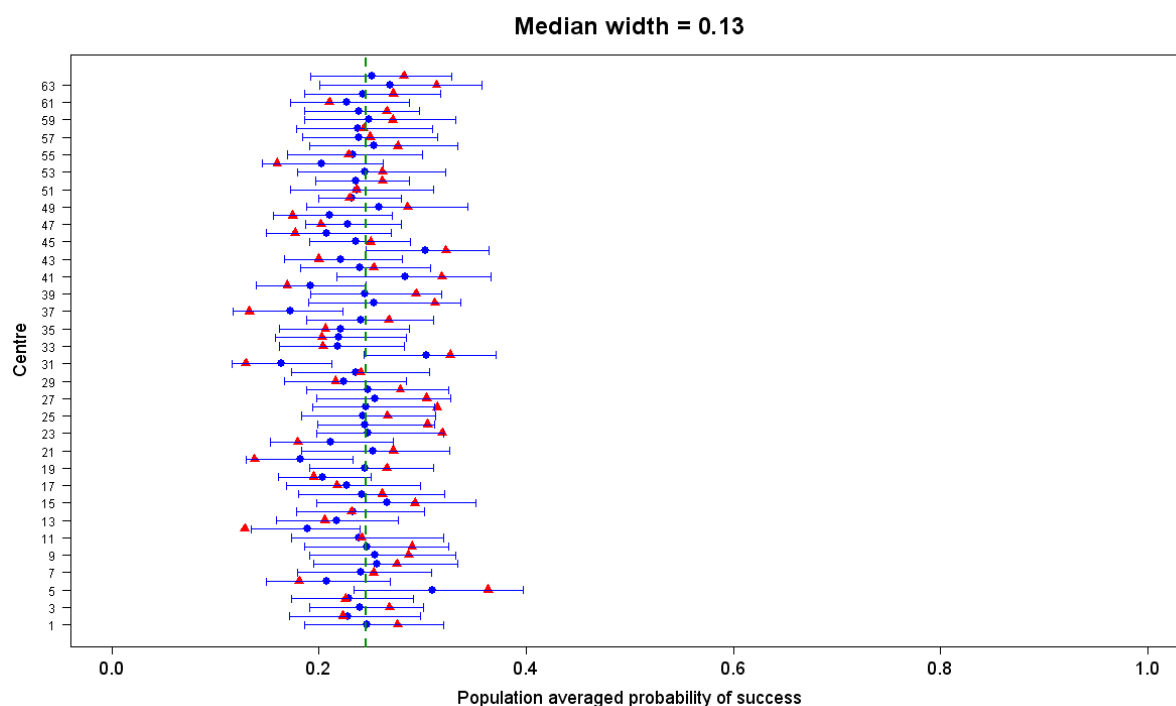


Figure 33: On QCI 1232a (version 2): population averaged probability of success for the **hierarchical logistic regression** model, per centre. The red triangles represent the 'true' population averaged probabilities of success, the blue bullets represent the average population averaged probabilities of success over 1000 simulations and the intervals are based on the empirical distribution of all simulated population averaged probabilities of success.

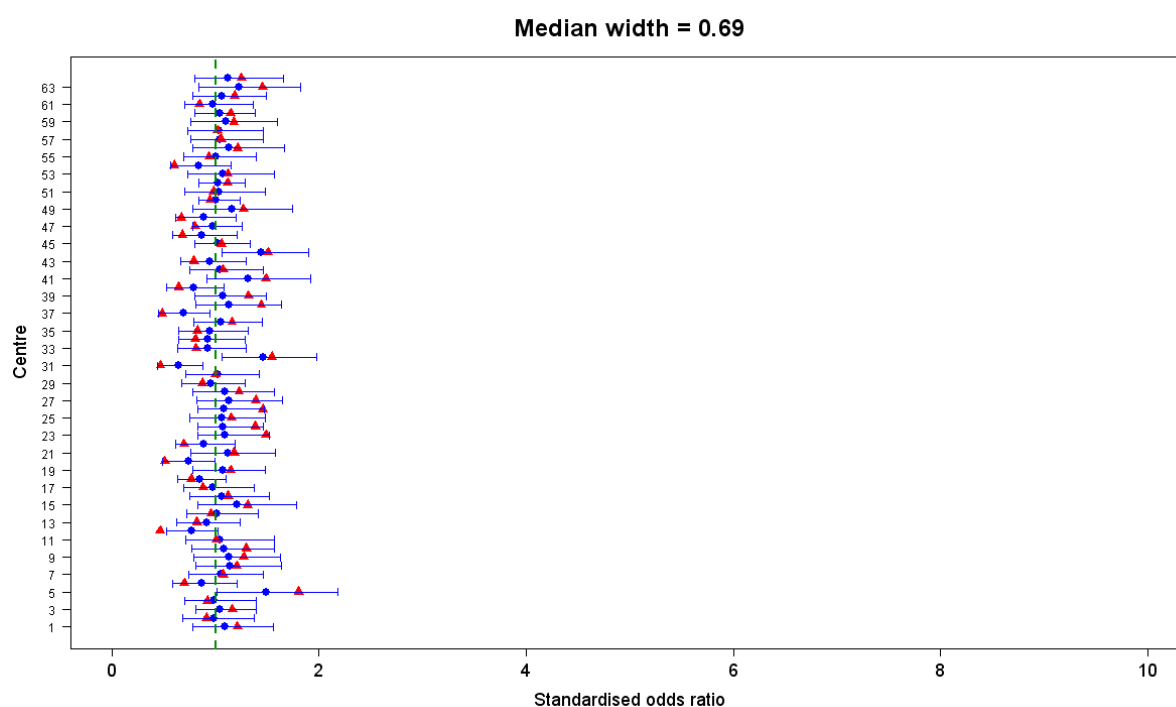


Figure 34: On QCI 1232a (version 2): (centre-specific) standardised odds ratio for the **hierarchical logistic regression** model, per centre. The red triangles represent the 'true' standardised odds ratio's, the blue bullets represent the average standardised odds ratio's over 1000 simulations and the intervals are based on the empirical distribution of all simulated odds ratio's.

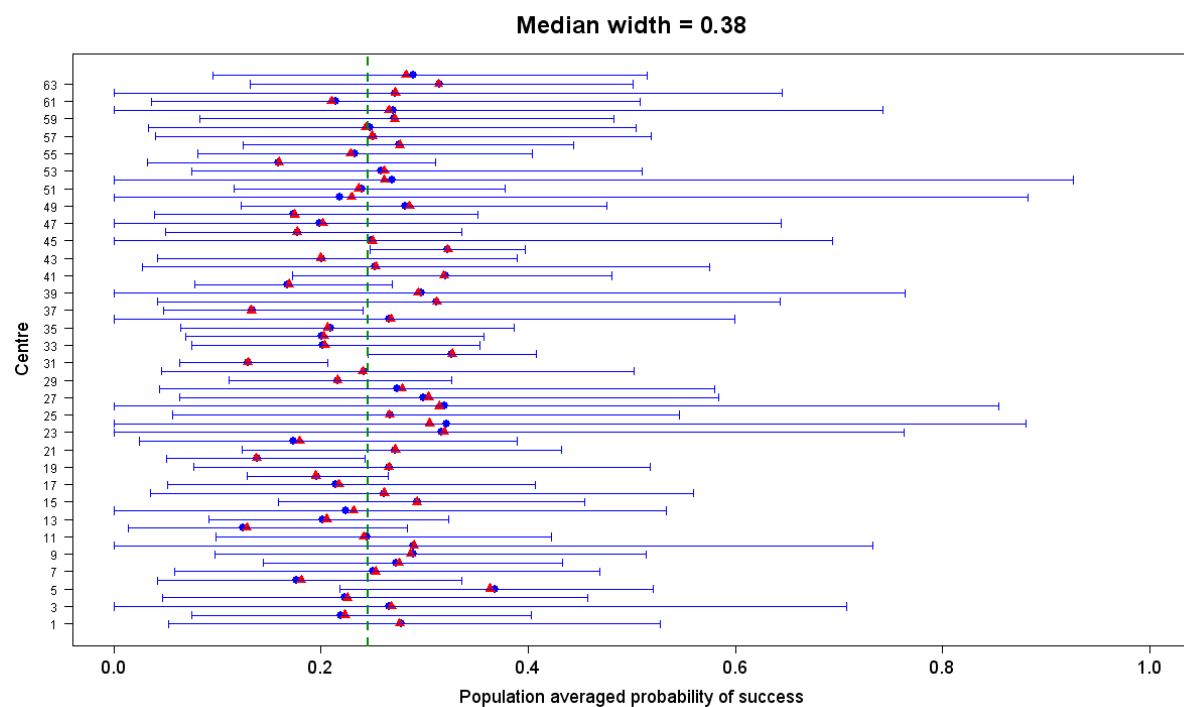


Figure 35: On QCI 1232a (version 2): population averaged probability of success for the **propensity score method**, per centre. The red triangles represent the 'true' population averaged probabilities of success, the blue bullets represent the average population averaged probabilities of success over 1000 simulations and the intervals are based on the empirical distribution of all simulated population averaged probabilities of success.

C Evaluation of simulations for QCI 1111, version 1

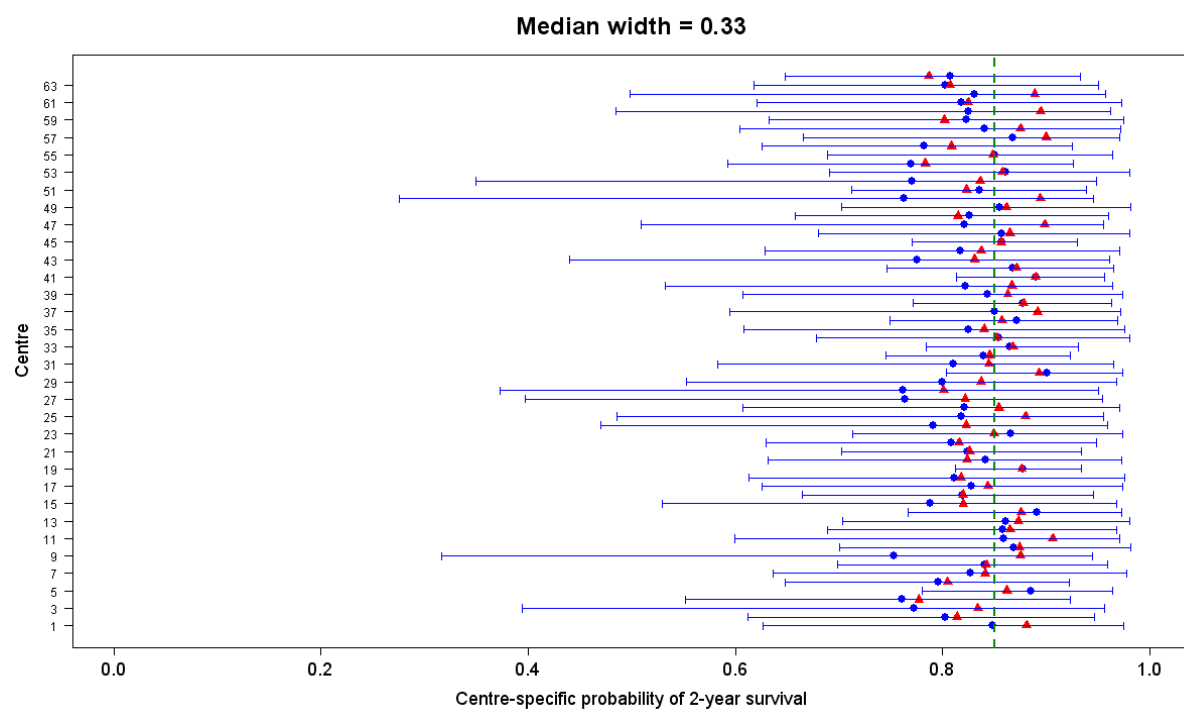


Figure 36: On QCI 1111 (version 1): centre-specific probabilities of 2-year survival for the **(Firth corrected) fixed effects Cox proportional hazards** model, per centre. The red triangles represent the 'true' centre-specific probabilities of 2-year survival, the blue bullets represent the average centre-specific probabilities of 2-year survival over 1000 simulations and the intervals are based on the empirical distribution of all centre-specific probabilities of 2-year survival.

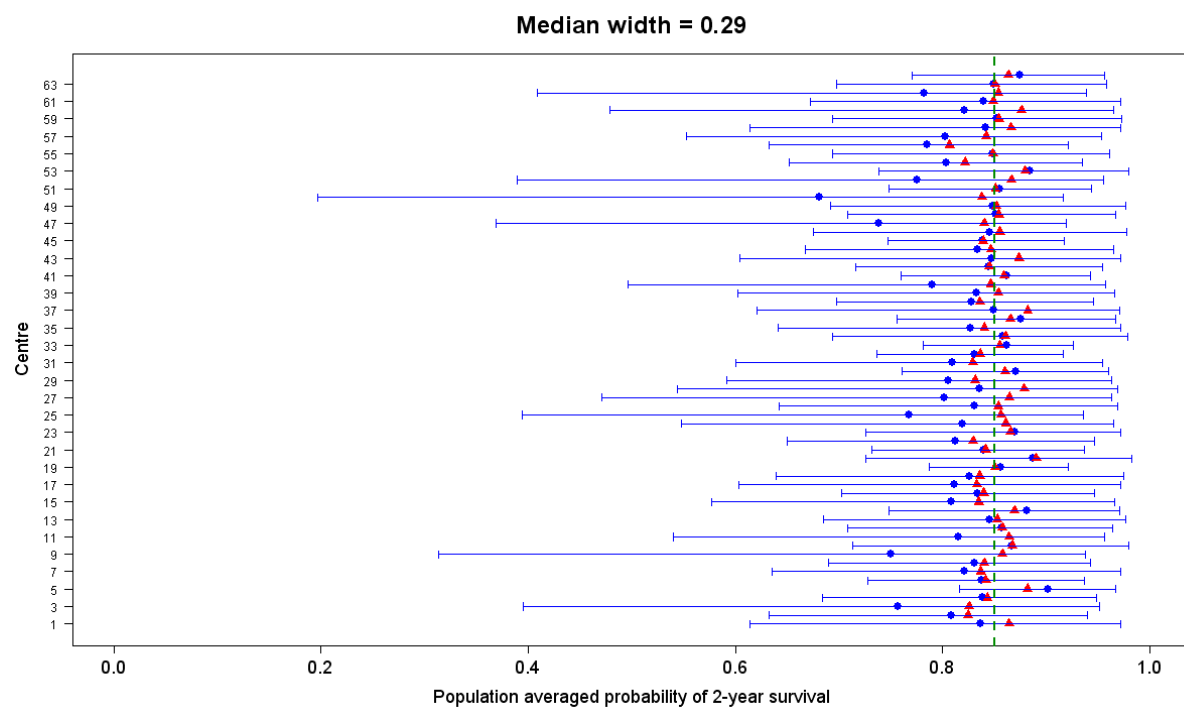


Figure 37: On QCI 1111 (version 1): population averaged probabilities of 2-year survival for the **(Firth corrected) fixed effects Cox proportional hazards** model, per centre. The red triangles represent the 'true' population averaged probabilities of 2-year survival, the blue bullets represent the average population averaged probabilities of 2-year survival over 1000 simulations and the intervals are based on the empirical distribution of all population averaged probabilities of 2-year survival.

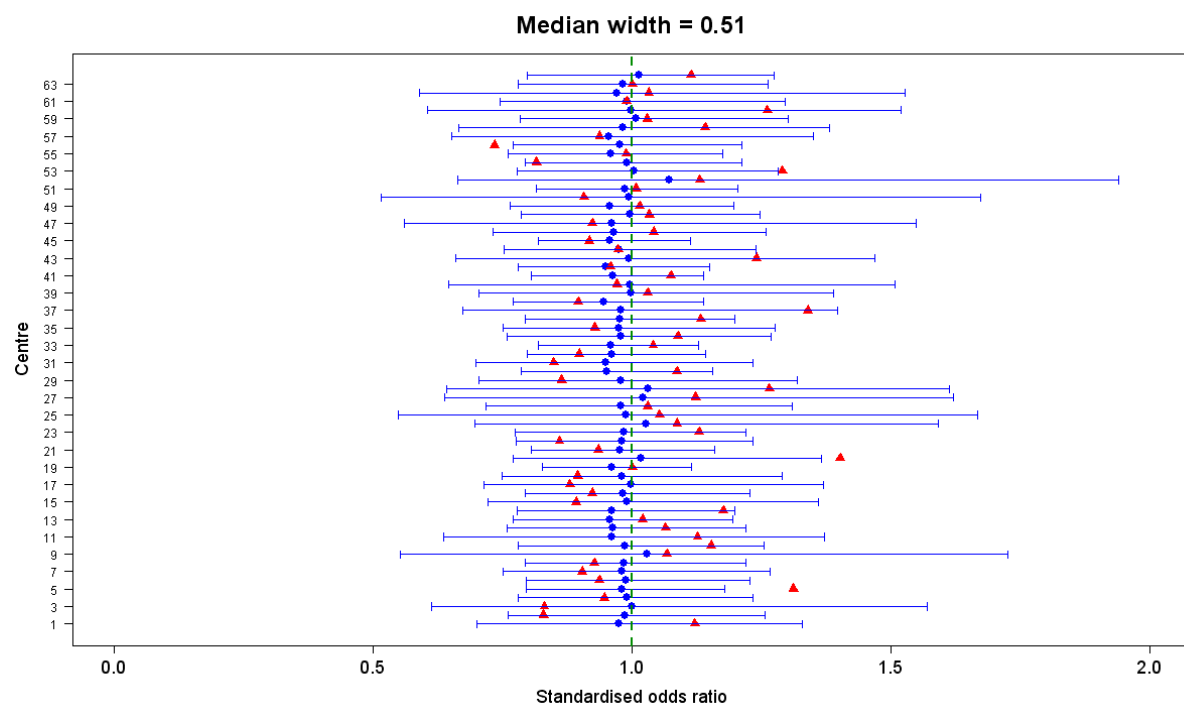


Figure 38: On QCI 1111 (version 1): (centre-specific) standardised odds ratio of 2-year survival for the **(Firth corrected) fixed effects Cox proportional hazards** model, per centre. The red triangles represent the 'true' standardised odds ratio's of 2-year survival, the blue bullets represent the average standardised odds ratio's of 2-year survival over 1000 simulations and the intervals are based on the empirical distribution of all standardised odds ratio's of 2-year survival.

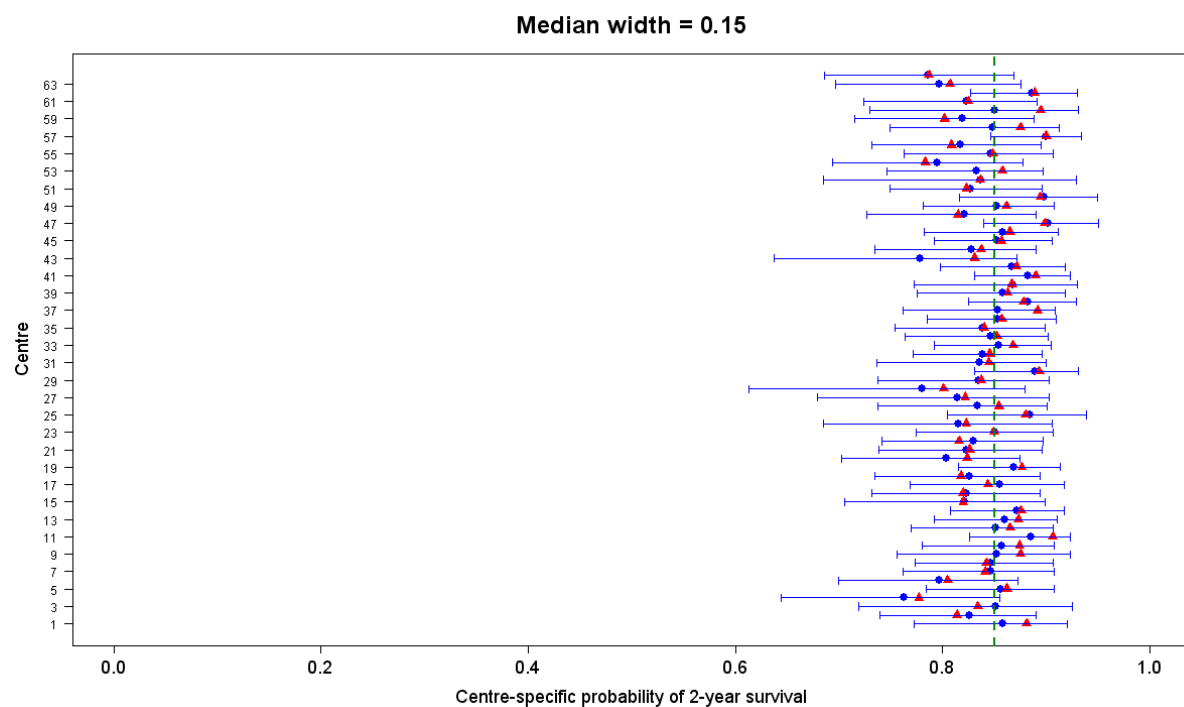


Figure 39: On QCI 1111 (version 1): centre-specific probabilities of 2-year survival for the **frailty Cox proportional hazards** model, per centre. The red triangles represent the 'true' centre-specific probabilities of 2-year survival, the blue bullets represent the average centre-specific probabilities of 2-year survival over 1000 simulations and the intervals are based on the empirical distribution of all centre-specific probabilities of 2-year survival.

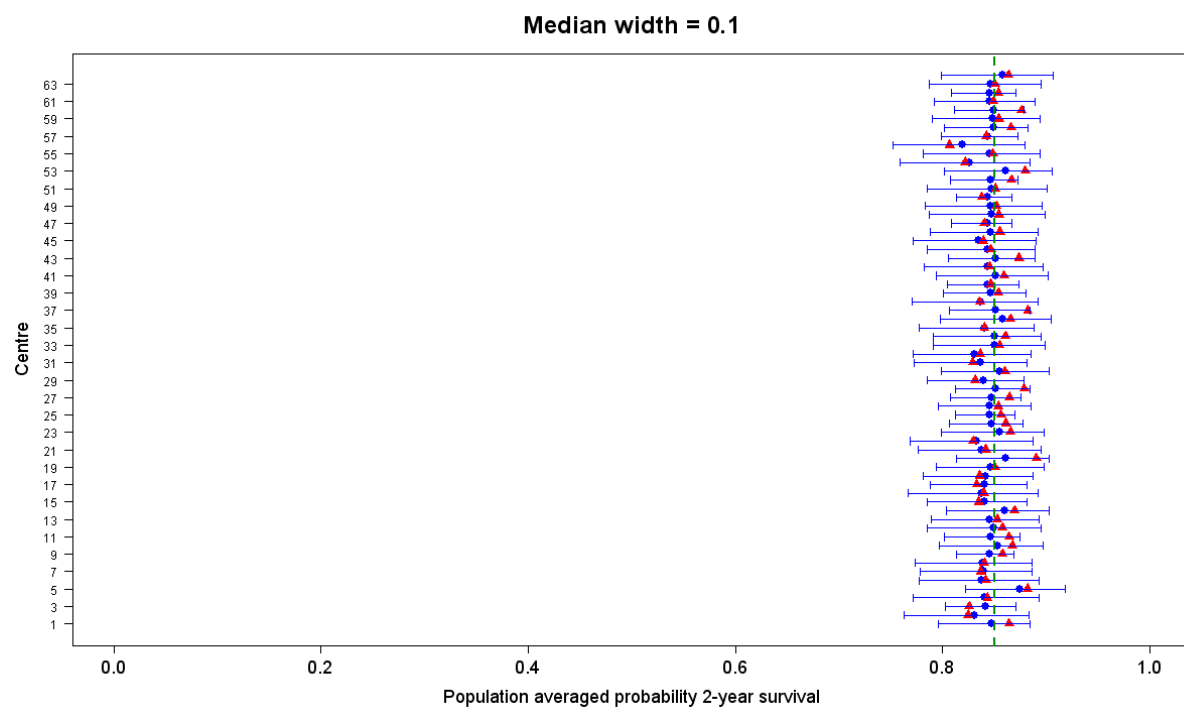


Figure 40: On QCI 1111 (version 1): population averaged probabilities of 2-year survival for the **frailty Cox proportional hazards** model, per centre. The red triangles represent the 'true' population averaged probabilities of 2-year survival, the blue bullets represent the average population averaged probabilities of 2-year survival over 1000 simulations and the intervals are based on the empirical distribution of all population averaged probabilities of 2-year survival.

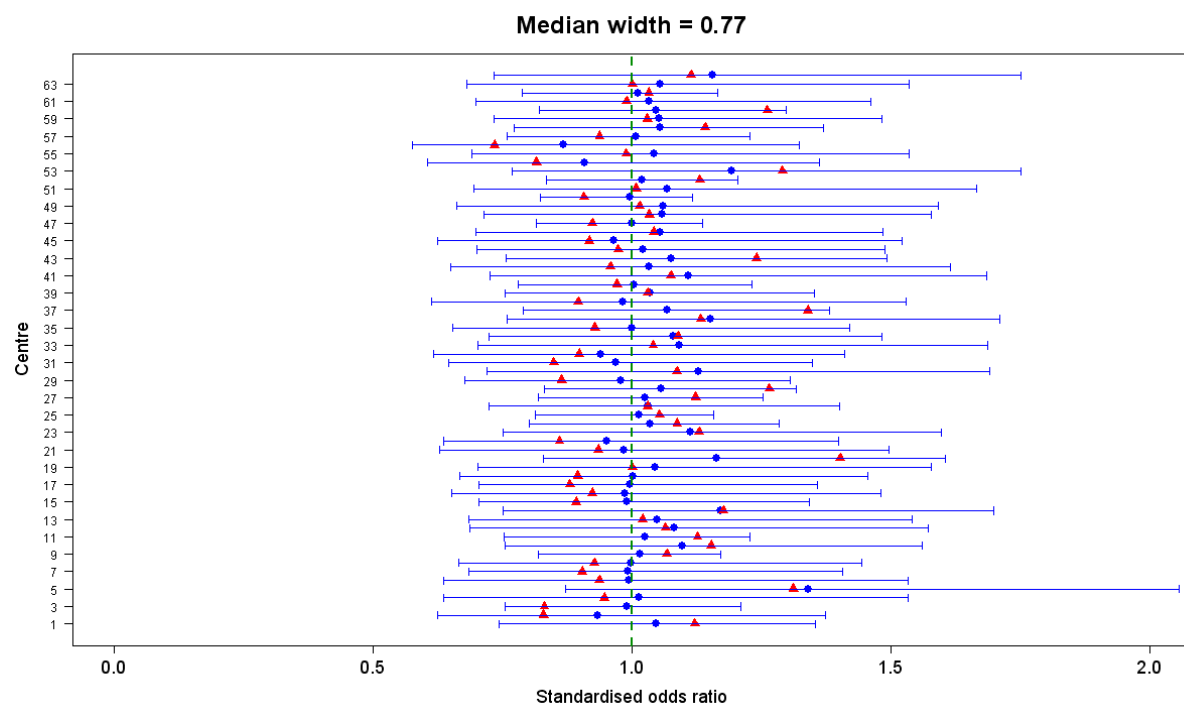


Figure 41: On QCI 1111 (version 1): (centre-specific) standardised odds ratio of 2-year survival for the **frailty Cox proportional hazards** model, per centre. The red triangles represent the 'true' standardised odds ratio's of 2-year survival, the blue bullets represent the average standardised odds ratio's of 2-year survival over 1000 simulations and the intervals are based on the empirical distribution of all standardised odds ratio's of 2-year survival.

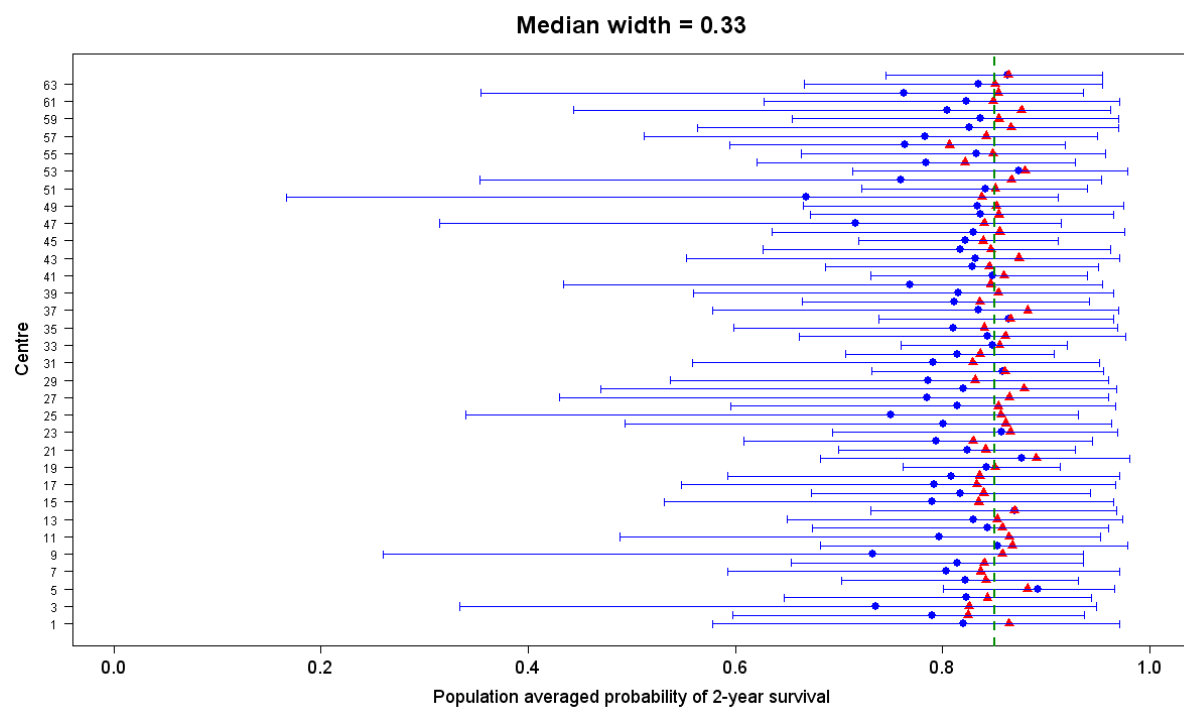


Figure 42: On QCI 1111 (version 1): population averaged probabilities of 2-year survival for the **propensity score method**, per centre. The red triangles represent the 'true' population averaged probabilities of 2-year survival, the blue bullets represent the average population averaged probabilities of 2-year survival over 1000 simulations and the intervals are based on the empirical distribution of all population averaged probabilities of 2-year survival.

Contents

1	Estimation of center effects	2
1.1	Binary QCI	2
1.1.1	Logistic regression model (fixed-effects)	2
1.1.2	Hierarchical logistic regression model	5
1.1.3	Doubly-robust propensity score method	5
1.2	Right-censored QCI	7
1.2.1	Cox' proportional hazards model	7
1.2.2	Cox' frailty proportional hazards model	8
1.3	Continuous QCI - Linear regression model	9
2	All or none (outcome) quality index	10

1 Estimation of center effects

Similarly as in Normand et al. (1997) centers will be evaluated based on a comparison between the average expected outcome for the patients they treated and the average expected outcome their patients would have if they were to be treated at the 'average' center. We will call this measure the center-averaged excess outcome and denote it with e_c , it is obtained in a different way depending on the type of QCI. Center effects will be expressed as an 'excess' probability or outcome value, i.e. the obtained center-specific mean will be 'standardized' by subtracting the probability or mean outcome value one would expect to observe if all patients of that specific center had been treated in the average center. These 'excess' probabilities or outcome values are computed in a different way, depending on the type of QCI: binary, right-censored or continuous.

To determine error bars for the unadjusted and adjusted caterpillar plots, we first need an estimate for $\text{Var}(\hat{e}_c)$. This variance is obtained through different methods, depending on the type of QCI: for continuous QCIs the variance is estimated directly in the model, the (asymptotic) Delta-method is used for binary QCIs and a bootstrap procedure for right-censored QCIs.

Before embarking in any analysis, for each QCI, centers with less than five patients eligible patients for that QCI are merged into one overlapping center and further analysed as if all these patients were treated in that 'overlapping/merged' center. Obviously, since different patients can be eligible for different QCIs, this overlapping center does not always consists of the same centers, hence interpretations in this regard should be made carefully.

1.1 Binary QCI

1.1.1 Logistic regression model (fixed-effects)

For binary QCIs a logistic regression model with the QCI as outcome and identified prognostic factors and (effect-coded) center choice (with ψ_c the estimated effect for center c) as predictors:

$$\text{logit}(E(Y_{ic}|\mathbf{X}_i, \mathbf{L}_i)) = \alpha + \mathbf{L}_i' \boldsymbol{\beta} + \sum_{c \neq \text{ref}} \psi_c X_{ic},$$

with X_{ic} the effect-coded center 'indicators' (for patient i treated in center c with $c \neq \text{ref}$):

c	X_{i1}	X_{i2}	X_{i3}	...	$X_{i,m}$
1	1	0	0	...	0
2	0	1	0	...	0
3	0	0	1	...	0
\vdots					
ref	-1	-1	-1	...	-1
\vdots					
m	0	0	0	...	1

Note that the reference center was chosen as the centers with overall most entries in the PROCARE database.

The center-averaged excess probability is obtained from the fitted coefficients in this model as follows:

$$\hat{e}_c = \frac{1}{n_c} \sum_{i=1}^{n_c} \left(\text{expit}(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta} + \hat{\psi}_c) - \text{expit}(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta}) \right), \quad (1)$$

with $\hat{\psi}_{ref} = -\sum_{c \neq ref} \hat{\psi}_c$.

Determination of error bars The variance of \hat{e}_c can be decomposed as follows:

$$\begin{aligned} \text{Var}(\hat{e}_c) &= \text{Var} \left(\frac{1}{n_c} \sum_{i=1}^{n_c} \left(\text{expit}(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta} + \hat{\psi}_c) - \text{expit}(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta}) \right) \right) \\ &= \frac{1}{n_c^2} \text{Var} \left(\sum_{i=1}^{n_c} \text{expit}(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta} + \hat{\psi}_c) - \sum_{i=1}^{n_c} \text{expit}(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta}) \right) \\ &= \frac{1}{n_c^2} \left[\text{Var} \left(\sum_{i=1}^{n_c} \text{expit}(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta} + \hat{\psi}_c) \right) + \text{Var} \left(\sum_{i=1}^{n_c} \text{expit}(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta}) \right) \right] \\ &\quad - \frac{2}{n_c^2} \text{Cov} \left(\sum_{i=1}^{n_c} \text{expit}(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta} + \hat{\psi}_c), \sum_{i=1}^{n_c} \text{expit}(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta}) \right) \end{aligned}$$

For each of the three terms in the latter equation, the **Delta-method** needs to be applied, we will only show this for the first term: $\text{Var} \left(\sum_{i=1}^{n_c} \text{expit}(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta} + \hat{\psi}_c) \right)$. Analogous steps need to be followed to obtain $\text{Var} \left(\sum_{i=1}^{n_c} \text{expit}(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta}) \right)$ and $\text{Cov} \left(\sum_{i=1}^{n_c} \text{expit}(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta} + \hat{\psi}_c), \sum_{i=1}^{n_c} \text{expit}(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta}) \right)$.

- STEP 1: Suppose

$$\begin{pmatrix} \hat{\alpha} \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_I \\ \hat{\psi}_c \end{pmatrix} \sim N \left[\begin{pmatrix} \alpha \\ \beta_1 \\ \vdots \\ \beta_I \\ \psi_c \end{pmatrix}, \Sigma_{|\alpha, \beta_1, \dots, \beta_I, \psi_c} \right],$$

where $\Sigma_{|\alpha, \beta_1, \dots, \beta_I, \psi_c}$ is the variance-covariance matrix of the fitted coefficients restricted to the parameter estimates $\alpha, \beta_1, \dots, \beta_I$ and ψ_c .

- STEP 2: Define the function

$$f(\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_I, \hat{\psi}_c | \mathbf{L}_{ic}) = \sum_{i=1}^{n_c} \text{expit}(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta} + \hat{\psi}_c).$$

- STEP 3: Compute the partial derivatives with respect to all the parameters in the function $f(\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_I, \hat{\psi}_c | \mathbf{L}_{ic})$:

$$\frac{\partial f(\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_I, \hat{\psi}_c | \mathbf{L}_{ic})}{\partial \hat{\alpha}} = \sum_{i=1}^{n_c} \frac{\exp(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta} + \hat{\psi}_c)(1 + \exp(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta} + \hat{\psi}_c)) - \exp(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta} + \hat{\psi}_c)^2}{(1 + \exp(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta} + \hat{\psi}_c))^2},$$

$$\frac{\partial f(\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_I, \hat{\psi}_c | \mathbf{L}_{ic})}{\partial \hat{\beta}_1} = \sum_{i=1}^{n_c} \frac{\exp(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta} + \hat{\psi}_c)(1 + \exp(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta} + \hat{\psi}_c)) - \exp(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta} + \hat{\psi}_c)^2}{(1 + \exp(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\beta} + \hat{\psi}_c))^2} L_{1ic},$$

⋮

$$\frac{\partial f(\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_I, \hat{\psi}_c | \mathbf{L}_{ic})}{\partial \hat{\beta}_I} = \sum_{i=1}^{n_c} \frac{\exp(\hat{\alpha} + \mathbf{L}'_{ic} \hat{\beta} + \hat{\psi}_c)(1 + \exp(\hat{\alpha} + \mathbf{L}'_{ic} \hat{\beta} + \hat{\psi}_c)) - \exp(\hat{\alpha} + \mathbf{L}'_{ic} \hat{\beta} + \hat{\psi}_c)^2}{(1 + \exp(\hat{\alpha} + \mathbf{L}'_{ic} \hat{\beta} + \hat{\psi}_c))^2} L_{lic},$$

$$\frac{\partial f(\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_I, \hat{\psi}_c | \mathbf{L}_{ic})}{\partial \hat{\psi}_c} = \sum_{i=1}^{n_c} \frac{\exp(\hat{\alpha} + \mathbf{L}'_{ic} \hat{\beta} + \hat{\psi}_c)(1 + \exp(\hat{\alpha} + \mathbf{L}'_{ic} \hat{\beta} + \hat{\psi}_c)) - \exp(\hat{\alpha} + \mathbf{L}'_{ic} \hat{\beta} + \hat{\psi}_c)^2}{(1 + \exp(\hat{\alpha} + \mathbf{L}'_{ic} \hat{\beta} + \hat{\psi}_c))^2}.$$

- STEP 4: Define the vector

$$\frac{\partial \mathbf{f}}{\partial} = \begin{pmatrix} \frac{\partial f(\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_I, \hat{\psi}_c | \mathbf{L}_{ic})}{\partial \hat{\alpha}} \\ \frac{\partial f(\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_I, \hat{\psi}_c | \mathbf{L}_{ic})}{\partial \hat{\beta}_1} \\ \vdots \\ \frac{\partial f(\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_I, \hat{\psi}_c | \mathbf{L}_{ic})}{\partial \hat{\beta}_I} \\ \frac{\partial f(\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_I, \hat{\psi}_c | \mathbf{L}_{ic})}{\partial \hat{\psi}_c} \end{pmatrix}$$

- STEP 5: Estimate

$$\widehat{\text{Var}} \left(\sum_{i=1}^{n_c} \text{expit}(\hat{\alpha} + \mathbf{L}'_{ic} \hat{\beta} + \hat{\psi}_c) \right) = \frac{\partial \mathbf{f}'}{\partial} \hat{\Sigma}_{|\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_I, \hat{\psi}_c} \frac{\partial \mathbf{f}}{\partial},$$

where $\hat{\Sigma}_{|\hat{\alpha}, \hat{\beta}_1, \dots, \hat{\beta}_I, \hat{\psi}_c}$ is the estimated version of $\Sigma_{|\alpha, \beta_1, \dots, \beta_I, \psi_c}$.

Finally, the error bars that are used in the caterpillar plot, are obtained as:

$$\left[\hat{e}_c \pm t_{n_c-1, 0.975} \sqrt{\widehat{\text{Var}}(\hat{e}_c)} \right].$$

Note Variance- and covariance for the reference center

The center-effect for the reference center (say $c = \text{ref}$) is obtained with $\hat{\psi}_{\text{ref}} = -(\hat{\psi}_1 + \hat{\psi}_2 + \dots + \hat{\psi}_{m-1})$, hence

$$\begin{aligned} \widehat{\text{Var}}(\hat{\psi}_{\text{ref}}) &= (-1 - 1 \dots - 1) \hat{\Sigma}_{|\hat{\psi}_1, \hat{\psi}_2, \dots, \hat{\psi}_{m-1}} \begin{pmatrix} -1 \\ -1 \\ \vdots \\ -1 \end{pmatrix} \\ \widehat{\text{Cov}}(\hat{\psi}_{\text{ref}}, \hat{\alpha}) &= (0 - 1 - 1 \dots - 1) \hat{\Sigma}_{|\hat{\alpha}, \hat{\psi}_1, \hat{\psi}_2, \dots, \hat{\psi}_{m-1}} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\ \widehat{\text{Cov}}(\hat{\psi}_{\text{ref}}, \hat{\beta}_1) &= (0 - 1 - 1 \dots - 1) \hat{\Sigma}_{|\hat{\beta}_1, \hat{\psi}_1, \hat{\psi}_2, \dots, \hat{\psi}_{m-1}} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \\ &\vdots \\ \widehat{\text{Cov}}(\hat{\psi}_{\text{ref}}, \hat{\beta}_I) &= (0 - 1 - 1 \dots - 1) \hat{\Sigma}_{|\hat{\beta}_I, \hat{\psi}_1, \hat{\psi}_2, \dots, \hat{\psi}_{m-1}} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \end{aligned}$$

1.1.2 Hierarchical logistic regression model

A hierarchical logistic regression model with a patient- and center level can be presented as follows:

- A patient-level logistic regression model:

$$\text{logit}(E(Y_{ic}|\mathbf{L}_i)) = \alpha_c + \mathbf{L}_i' \boldsymbol{\beta}.$$

- A center-level linear regression model:

$$\alpha_c = \alpha + \psi_c, \quad \psi_c \sim N(0, \sigma_\psi^2)$$

Combining the two level model yields

$$\text{logit}(E(Y_{ic}|\mathbf{L}_i)) = \alpha + \mathbf{L}_i' \boldsymbol{\beta} + \psi_c, \quad \psi_c \sim N(0, \sigma_\psi^2).$$

The center-averaged excess probability is obtained from the fitted coefficients and *best linear unbiased predictors* (BLUPs), $\hat{\psi}_c$, in this model exactly as in (1).

Note The standard modelling procedure in SAS does not always converge, especially in cases where there are many centers with few patients and no or only events. In most cases this issue could be resolved by using a different optimization algorithm (Laplace optimization). If this procedure still does not produce reliable results for a certain QCI, no excess probabilities were computed for this method.

1.1.3 Doubly-robust propensity score method

Here, the focus is on estimating the 'excess' probability of success that would have been observed had all patients in the study population been treated at center c , relative to this probability in the average center. Let $Y_i(c)$ denote the probability of 'success' that would have been observed for a given patient i if treated at center c . Then more formally our interest lies in estimating the counterfactual probability $P\{Y(c) = 1\} = E\{Y_i(c)\}$.

The *outcome working model* is a logistic regression model for the probability of 'success' in center c in function of prognostic factors, i.e.

$$m(\mathbf{L}; \alpha, \boldsymbol{\beta}, \boldsymbol{\psi}) = \text{logit}(E(Y_i|\mathbf{X}_i, \mathbf{L}_i)) = \alpha + \mathbf{L}_i' \boldsymbol{\beta} + \sum_{c \neq \text{ref}} \psi_c X_{ic},$$

with X_{ic} the effect-coded center 'indicators'.

The *propensity score working model* is a multinomial regression model for the multiple propensity score: the probability of a patient being treated in center c in function of prognostic factors, i.e.

$$h(c, \mathbf{L}; \alpha_c^*, \boldsymbol{\beta}_c^*) = P(I(\mathbf{X}_i = c)|\mathbf{L}_i) = \begin{cases} \frac{1}{1 + \sum_{j \neq \text{ref}} \exp(\alpha_j^* + \mathbf{L}_j' \boldsymbol{\beta}_j^*)} & c = \text{ref} \\ \frac{\exp(\alpha_c^* + \mathbf{L}_c' \boldsymbol{\beta}_c^*)}{1 + \sum_{j \neq \text{ref}} \exp(\alpha_j^* + \mathbf{L}_j' \boldsymbol{\beta}_j^*)} & c \neq \text{ref} \end{cases}.$$

By fitting the outcome working model via a weighted regression of outcome on covariates amongst patients attending center c , with weights w_i , determined as follows (Robins et al. (2008)):

$$w_i^{(1)} = \frac{n_{ic}}{m} \frac{1}{h(c, \mathbf{L}_i; \hat{\alpha}_c^*, \hat{\beta}_c^*)},$$

which is truncated at its own 1%- and 99%-tiles:

$$w_i = \begin{cases} p_{1\%}^{w_i^{(1)}} & w_i^{(1)} \leq p_{1\%}^{w_i^{(1)}} \\ w_i^{(1)} & p_{1\%}^{w_i^{(1)}} \leq w_i^{(1)} \leq p_{99\%}^{w_i^{(1)}} \\ p_{99\%}^{w_i^{(1)}} & w_i^{(1)} \geq p_{99\%}^{w_i^{(1)}} \end{cases}$$

The doubly robust estimator can then be obtained as

$$E\{\widehat{Y_i(c)}\} = \frac{1}{n} \sum_{i=1}^n m(c, \mathbf{L}_i; \hat{\alpha}, \hat{\beta}, \hat{\psi}_c).$$

Note that centers with no events or only events are not entered in the analysis, but instead their counterfactual probability is automatically set to 0 or 1, respectively.

A population averaged variant of the excess probability is then obtained as

$$\hat{e}_c^{\text{pop}} = \begin{cases} -\frac{1}{n} \sum_{i=1}^n m(c, \mathbf{L}_i; \hat{\alpha}, \hat{\beta}, 0) & \sum_{i=1}^{n_c} Y_{ic} = 0 \\ \frac{1}{n} \sum_{i=1}^n \left(m(c, \mathbf{L}_i; \hat{\alpha}, \hat{\beta}, \hat{\psi}_c) - m(c, \mathbf{L}_i; \hat{\alpha}, \hat{\beta}, 0) \right) & \sum_{i=1}^{n_c} Y_{ic} \neq (0, n_c) \\ 1 - \frac{1}{n} \sum_{i=1}^n m(c, \mathbf{L}_i; \hat{\alpha}, \hat{\beta}, 0) & \sum_{i=1}^{n_c} Y_{ic} = n_c \end{cases}.$$

Apart from simplicity, an attraction of this estimator is that, unlike the other doubly robust estimators, it does not inflate the bias due to model misspecification in regions where the weights are large (Vansteelandt et al. (2010)). A further attraction is that, because the estimator is calculated as the average of the fitted values from a binary outcome regression model, it guarantees probability estimates within the $[0 - 1]$ interval.

As mentioned in the main report of Deliverable 6, the small size of various centers in the PROCARE setting does not enable accurate assessment of the propensity score working model, $h(c, \mathbf{L}_i; \alpha_c^*, \beta_c^*)$. Future research will examine more closely a number of strategies that may help improve the performance of propensity score methods when the number of patients per center is relatively small. We foresee the largest benefit via penalization methods that shrink the regression coefficients in the multinomial propensity score model towards zero. This will prevent variance inflation and thereby induce greater stability in the inverse propensity score weights. An alternative strategy would be based on stabilized estimation of the propensity score in a way that targets maximal precision of the final analysis results. Such strategy is for instance proposed in Vansteelandt et al. (2010), where it is found to result in much greater stability of the inverse propensity score weights and, thereby, in more precise analysis results. Combinations of both strategies whereby the shrinkage parameters are chosen to minimize (mean squared) error in the final analysis results, may also be considered if computationally feasible.

1.2 Right-censored QCI

1.2.1 Cox' proportional hazards model

For right-censored QCIs a Firth-corrected (cStage-stratified) Cox proportional hazards model with the QCI as outcome and identified prognostic factors and center choice as predictors:

$$\lambda_i(t, \mathbf{X}_{ic}, \mathbf{L}_i) = \lambda_{0,s_i}(t) \exp \left(\mathbf{L}'_i \boldsymbol{\beta} + \sum_{c \neq \text{ref}} \psi_c \mathbf{X}_{ic} \right), \quad (2)$$

with \mathbf{X}_{ic} the effect-coded center 'indicators'.

The center-averaged 'excess' probability of x -year survival is obtained from the fitted coefficients as follows:

$$\hat{e}_c(x) = \frac{1}{n_c} \sum_{i=1}^{n_c} (\hat{S}_{0,s_i}(x)^{\exp(\mathbf{L}'_i \hat{\boldsymbol{\beta}} + \hat{\psi}_c)} - \hat{S}_{0,s_i}(x)^{\exp(\mathbf{L}'_i \hat{\boldsymbol{\beta}})}), \quad (3)$$

with $\hat{\psi}_{\text{ref}} = -\sum_{c \neq \text{ref}} \hat{\psi}_c$.

Competing risks setting In a competing risks setting, we evaluate the centers in terms of their performance for the various possible outcomes. The evaluation is based on fitting Firth-corrected (cStage-stratified) cause-specific proportional hazard models, including identified prognostic factors and center choice as predictors. With f the failure type under investigation ($f \in \{0, 1\}$), this model becomes:

$$\lambda_{f,i}(t, \mathbf{X}_{ic}, \mathbf{L}_{f,i}) = \lambda_{f,0,s_i}(t) \exp \left(\mathbf{L}'_{f,i} \boldsymbol{\beta}_f + \sum_{c \neq \text{ref}} \psi_{f,c} \mathbf{X}_{ic} \right),$$

with \mathbf{X}_{ic} the effect-coded center 'indicators'.

While the center choice indicators, \mathbf{X}_{ic} , are fixed in these models, the set of prognostic factors \mathbf{L} is determined through significance tests, and differs between the target causes (yielding a set of covariates \mathbf{L}_f with coefficients $\boldsymbol{\beta}_f$). Also, the centers have different effects for the different causes (yielding a set of coefficients $\psi_{f,c}$).

Once the cause-specific models are fit, they are combined to yield - for each patient - the *cumulative incidence* for the cause of interest ($f = 1$), i.e. the a priori probability of seeing the event of interest before time t , taking the competing risk ($f = 0$) into account:

$$\begin{aligned} \Lambda_{1,i}(t; \mathbf{X}_{ic}, \mathbf{L}_{0,i}, \mathbf{L}_{1,i}) &= \int_0^t \lambda_{1,i}(u, \mathbf{X}_{ic}, \mathbf{L}_{1,i}) S_i(u) du \\ &= \int_0^t \lambda_{1,0,s_i}(u) \exp \left(\mathbf{L}'_{1,i} \boldsymbol{\beta}_1 + \sum_{c \neq \text{ref}} \psi_{1,c} \mathbf{X}_{ic} \right) \times \\ &\quad S_{1,0,s_i}(u)^{\exp(\mathbf{L}'_{1,i} \boldsymbol{\beta}_1 + \sum_{c \neq \text{ref}} \psi_{1,c} \mathbf{X}_{ic})} S_{0,0,s_i}(u)^{\exp(\mathbf{L}'_{0,i} \boldsymbol{\beta}_0 + \sum_{c \neq \text{ref}} \psi_{0,c} \mathbf{X}_{ic})} du \end{aligned}$$

The complement of this cumulative incidence is called the subsurvival function $S_1(t) = 1 - \Lambda_1(t)$, and yields the probability of not seeing an event of the specified type at time t . The final comparison of

centers is done in terms of the excess x -year subsurvival probability for the event of interest ($f = 1$) in each center c :

$$\hat{e}_{1,c}(x) = \frac{1}{n_c} \sum_{i=1}^{n_c} [S_{1,i}(x; \mathbf{X}_{ic}, \mathbf{L}_{0,i}, \mathbf{L}_{1,i}) - S_{1,i}(x; \mathbf{X}_{ic} = 0, \mathbf{L}_{0,i}, \mathbf{L}_{1,i})] \quad (4)$$

Note: while one can easily derive an overall excess probability in this setting, it will not be exactly equal to the standard survival excess defined in (3), since the underlying models differ.

Determination of error bars Error bars for caterpillar plots will be obtained from a bootstrap procedure, i.e. B random samples of the original PROCARE database are taken and in each random sample the center-specific excess x -year survival probabilities are computed. For each center, the 2.5%- and 97.5%-tile of the B obtained center effects are respectively the lower and upper limit for the error bars in the caterpillar plot.

Note that the bootstrap procedure for obtaining these error bars has not been fully developed till date and some further optimisation is needed before results will be presented.

1.2.2 Cox' frailty proportional hazards model

Frailty models are frequently used for modelling dependence in time-to-event data. They are the Cox-model equivalent of the random effects model. The aim of the frailty is to take the presence of correlation - due to some shared covariate information - between survival times into account, and correct for it without needing to fit a separate hazard (intercept) for each centre through a parametric model.

We will assume a *constant shared frailty*, or all patients i in centre c share the same frailty ψ_c . The variability of the ψ_c 's reflects the heterogeneity of risks between the m centres.

The conditional hazards model for patient i in stratum s in centre c is an extension of model (2)

$$\lambda_i(t, \psi) = \lambda_{0,s_i}(t) \exp(\mathbf{L}'_i \boldsymbol{\beta} + \psi_c) \quad (5)$$

The model assumes that all observed event-times are independent given the frailties, hence assumed a 'conditional independence' model.

The center-averaged 'excess' probability of x -year survival is obtained from the fitted coefficients as follows:

$$\hat{e}_c(x) = \frac{1}{n_c} \sum_{i=1}^{n_c} (\hat{S}_{0,s_i}(x)^{\exp(\mathbf{L}'_i \hat{\boldsymbol{\beta}} + \hat{\psi}_c)} - \hat{S}_{0,s_i}(x)^{\exp(\mathbf{L}'_i \hat{\boldsymbol{\beta}})}).$$

Competing risks setting Similarly as above, excess x -year survival probabilities can also be estimated in a competing risks setting by first computing the cumulative incidence function.

1.3 Continuous QCI - Linear regression model

For continuous QCIs the center-averaged excess outcome is obtained from the fitted coefficients in a linear regression model with the QCI as outcome and the patient-specific characteristics and center choice as predictors:

$$E(Y_{ic} | \mathbf{X}_i, \mathbf{L}_{ic}) = \alpha + \mathbf{L}'_{ic}\boldsymbol{\beta} + \sum_{c \neq ref} \psi_c X_{ic} + \epsilon_{ic},$$

with $\epsilon_{ic} \sim N(0, \sigma^2)$ and X_{ic} the effect-coded center 'indicators'.

The 'excess' outcome values are obtained as follows:

$$\hat{e}_c = \frac{1}{n_c} \sum_{i=1}^{n_c} \left(\hat{\alpha} + \mathbf{L}'_{ic}\hat{\boldsymbol{\beta}} + \hat{\psi}_c - (\hat{\alpha} + \mathbf{L}'_{ic}\hat{\boldsymbol{\beta}}) \right) = \hat{\psi}_c.$$

Determination of error bars $\text{Var}(\hat{e}_c) = \text{Var}(\hat{\psi}_c)$, hence can be extracted from the estimated variance-covariance matrix of the linear regression model.

For the reference center, with $\hat{\psi}_{ref} = -(\hat{\psi}_1 + \hat{\psi}_2 + \dots + \hat{\psi}_m)$, this variance is obtained as

$$\text{Var}(\hat{e}_{ref}) = (-1 - 1 \dots - 1) \hat{\Sigma}_{|\hat{\psi}_1, \hat{\psi}_2, \dots, \hat{\psi}_m} \begin{pmatrix} -1 \\ -1 \\ \vdots \\ -1 \end{pmatrix},$$

where $\hat{\Sigma}_{|\hat{\psi}_1, \hat{\psi}_2, \dots, \hat{\psi}_m}$ is the variance-covariance matrix of the fitted coefficients restricted to the parameter estimates $\hat{\psi}_1, \hat{\psi}_2, \dots, \hat{\psi}_m$.

2 All or none (outcome) quality index

The 'all or none' score for a patient indicates whether this patient reaches patient-level benchmarks for all QCIs for which it is eligible, in the case of the outcome quality index it involves QCI 1111 (Overall survival), QCI 1231 (Proportion of R0 resections), and QCI 1234b (Postoperative major surgical morbidity with reintervention under narcosis after radical surgical resection).

Translated, the all or none score indicates whether a patient reaches following benchmarks:

- whether he/she survived 3-years since incidence of rectal cancer,
- whether he/she had an R0 resection, and
- whether he/she did not have postoperative major surgical morbidity with reintervention under narcosis after radical surgical resection.

For the latter two indicators it is clear whether or not a (eligible) patient has reached the benchmark, but for patients with a follow-up of less than 3 years it is not that straightforward to determine whether they will survive 3 year or not. Therefore, a model-based multiple imputation technique is used to construct the all or none score and corresponding confidence limits, as described in the next steps.

Step 1 From the risk-adjustment model for QCI 1111, for all patients with less than 3-year follow-up who did not die within there observed follow-up period, the conditional probability of surviving 3 years after incidence of rectal cancer, given that they already survived up to the moment the administrative censoring date is computed.

Step 2 Using these conditional probabilities, 10 binary indicators are randomly determined for each of these patients.

Step 3 - repeated for each of the 10 indicators for QCI 1111 The all or none score is computed and center-specific excess probabilities ($\hat{e}_{c,k}$, with $k = 1, \dots, 10$), as well as the corresponding variance ($\widehat{\text{Var}}(\hat{e}_{c,k})$, obtained using the Delta-method) of these excess probabilities are estimated.

Step 4 The final center-specific excess probabilities are then obtained as

$$\hat{e}_{c,\text{all or none}} = \frac{1}{10} \sum_{k=1}^{10} \hat{e}_{c,k},$$

and the variance of this measure as

$$\widehat{\text{Var}}(\hat{e}_{c,\text{all or none}}) = \frac{1}{10} \sum_{k=1}^{10} \text{Var}(\hat{e}_{c,k}) + \left(1 + \frac{1}{10}\right) \frac{1}{9} \sum_{k=1}^{10} (\hat{e}_{c,k} - \hat{e}_{c,\text{all or none}})^2.$$

Step 5 Error bars for the caterpillar plot are then obtained as

$$\left[\hat{e}_{c, \text{all or none}} \pm t_{df, 0.975} \sqrt{\widehat{\text{Var}}(\hat{e}_{c, \text{all or none}})} \right],$$

with

$$df = (10 - 1) \left(1 + \frac{10 \hat{e}_{c, \text{all or none}}}{(10 + 1) \frac{1}{9} \sum_{k=1}^{10} (\hat{e}_{c, k} - \hat{e}_{c, \text{all or none}})^2} \right)^2.$$

References

- NORMAND, S. L. T., GLICKMAN, M. E. and GATSONIS, C. A. (1997). Statistical methods for profiling providers of medical care: Issues and applications. *Journal of the American Statistical Association*, **92** 803–814.
- ROBINS, J., SUED, M., LEI-GOMEZ, Q. and ROTNITZKY, A. (2008). Performance of double-robust estimators when inverse probability weights are highly variable. *Statistical Science*, **22** 544–559.
- VANSTEELANDT, S., BEKAERT, M. and CLAESKENS, G. (2010). On model selection and model misspecification in causal inference. *Statistical methods in medical research*, DOI: **10.1177/0962280210387717**.

Appendix 2: Protocol, results and discussion of the literature review

1	INTRODUCTION.....	130
2	SEARCH PROTOCOL.....	131
2.1	DETAILED SEARCH STRATEGY.....	132
2.2	PROGNOSTIC FACTORS	137
3	RESULTS	141
4	DISCUSSION.....	144
5	KEY POINTS	145
	REFERENCES.....	147

1 INTRODUCTION

The aim of the ProCare project is educational in the first place, i.e. individual centers receive feedback on the outcome of their rectal cancer patients as compared to all participating centers (all data of the entire Procure database are the benchmark). A fair comparison is only possible when the center's results are adjusted for all variables that may affect a patient's outcome irrespective of the therapy or therapies administered.

The statistical modeling approach that will be used to estimate the treatment center effect assumes that all confounders have been accounted for. When this is in doubt, a pseudo randomization approach can be used necessitating the availability of an instrumental variable, which predicts the choice of a treatment provider but does not by definition affect outcome. Therefore, the literature was searched for prognostic variables as well as instrumental variables in relation to rectal cancer.

2 SEARCH PROTOCOL

The following databases were searched: Medline through PubMed, Embase and the Cochrane Central Register of Controlled Trials. Details of the search strategy are presented below (section 2.1 detailed search strategy).

The identified articles were evaluated for relevance based on title and abstract by one person. Articles selected for full-text evaluation were divided among 5 researchers. Whenever available, confounding (prognostic) and/or instrumental variables were identified. Since none of the identified papers concerned interventional studies, no formal methodological assessment of the papers was performed. The selection criteria used are presented in table 1.

Table 1: Selection criteria used for full text evaluation.

Selection criteria	Inclusion criteria	Exclusion criteria
Population	colorectal or rectal cancer	subpopulations
Intervention	-	-
Outcome	patient specific factors with an influence on survival, recurrence, side effects, hospital choice of patient	respons to therapy (RT, CT, surgery), comparison between different therapies, diagnostic information (including screening), molecular information not related to prognosis, risk factors
Design	observational (prospective and retrospective) systematic reviews, meta-analysis	no multivariate analysis performed, interventional, randomized trial
Language	N, Fr, Eng, Ger	-

Included were studies that verified the prognostic significance of one or more independent clinical, pathological, or molecular variable(s) on outcome of rectal cancer using some form of multivariate analysis. Since the purpose of the search was to identify patient-specific (as opposed to treatment-related) confounding factors, studies reporting the effect of different therapeutic strategies tested in prospective randomized trials were excluded. Also, reports on epidemiological risk factors for (colo)rectal cancer were excluded.

2.1 DETAILED SEARCH STRATEGY

Author	Verhulst Johanna
Project number	2010-04_GCP (PPF2010-05_GCP)
Project name	Quality assurance of rectal cancer - phase 3: statistical methods to benchmark centres in a set of quality indicators
Search questions	Wich patient specific factors in patients with colorectal cancer can act as a confounder or instrumental variable on the quality of care indicators?
Keywords	rectal cancer, prognosis, mortality, recurrence

Date	July 13th 2010
Database (name + access; e.g.: Medline OVID)	Medline PubMed
Search Strategy (attention, for PubMed, check «Details»)	<p>#1 "Colorectal Neoplasms"[Mesh] OR "Rectal Neoplasms"[Mesh] OR ((colorectal[All Fields] OR ("rectum"[MeSH Terms] OR "rectum"[All Fields]) OR rectal[All Fields]) AND ((neoplasm* [All Fields] OR ("neoplasms"[MeSH Terms] OR "neoplasms"[All Fields] OR "cancer"[All Fields]) OR ("neoplasms"[MeSH Terms] OR "neoplasms"[All Fields] OR "cancers"[All Fields]) OR (carcino* [All Fields]) OR ("tumour"[All Fields] OR "neoplasms"[MeSH Terms] OR "neoplasms"[All Fields] OR "tumor"[All Fields]) OR ("tumour*" [All Fields] OR (metasta*[All Fields] OR (malign*[All Fields]))</p> <p>#3 Bibliography[Publication Type] OR Editorial[Publication Type] OR Letter[Publication Type] OR News[Publication Type]</p> <p>#7 ((prognostic score[All Fields] OR prognostic scores[All Fields] OR prognostic scoring[All Fields]) OR (prognostic[All Fields] AND ("abstracting and indexing as</p>

	<p>topic"[MeSH Terms] OR ("abstracting"[All Fields] AND "indexing"[All Fields] AND "topic"[All Fields]) OR "abstracting and indexing as topic"[All Fields] OR "index"[All Fields])) OR ("nomograms"[MeSH Terms] OR "nomograms"[All Fields] OR "nomogram"[All Fields]) OR (predictive[All Fields] AND model[All Fields]) OR validation[All Fields] OR validate[All Fields] OR (prognostic[All Fields] AND model[All Fields]) OR predictor[All Fields]) AND ((score*[All Fields])) OR ("abstracting and indexing as topic"[MeSH Terms] OR ("abstracting"[All Fields] AND "indexing"[All Fields] AND "topic"[All Fields]) OR "abstracting and indexing as topic"[All Fields] OR "index"[All Fields]) OR model[All Fields] OR (predict*[All Fields])</p> <p>OR (validat*[All Fields]) AND (multivar*[All Fields])</p> <p>((#7) AND #1) NOT #3 AND ("humans"[MeSH Terms] AND (English[lang] OR French[lang] OR German[lang] OR Dutch[lang]))</p>
Note	

Date	15 Jul 2010
Database (name + access; e.g.: Medline OVID)	Embase
Search Strategy (attention, for PubMed, check «Details»)	<p>#6. #4 AND #5</p> <p>#5. prognostic AND scor* OR prognostic AND index OR 'nomogram'/exp OR nomogram OR predictive AND (('model'/exp OR model) OR validation OR validate OR prognostic AND ('model'/exp OR model) OR predictor AND (score* OR scori* OR index OR 'model'/exp OR model OR predict* OR 'nomogram'/exp OR nomogram OR validat*) AND multivar*</p> <p>#4. #1 AND #2 AND #3</p> <p>#3. colorectal OR 'rectum'/exp OR rectum OR 'rectal'/exp OR rectal AND (neoplasm* OR 'cancer'/exp OR cancer OR 'cancers'/exp OR cancers OR carcinoma* OR carcinog* OR 'tumor'/exp OR tumor OR tumoral OR tumour* OR metastas* OR metastat* OR malign*)</p> <p>#2. 'rectum tumor'/exp AND [embase]/lim 7</p> <p>#1. 'colorectal tumor'/exp AND [embase]/lim</p>
Note	

Date	15 Jul 2010
Database (name + access; e.g.: Medline OVID)	Embase
Search Strategy (attention, for PubMed, check «Details»)	<p>#8. #4 AND #7</p> <p>#7. prognostic AND scor* OR prognostic AND index OR 'nomogram'/exp OR nomogram OR predictive AND (('model'/exp OR model) OR validation OR validate OR prognostic AND ('model'/exp OR model) OR predictor OR prognost* AND (score* OR scori* OR</p>

	<p>index OR 'model'/exp OR model OR predict* OR 'nomogram'/exp OR nomogram OR validat*) AND multivar*</p> <p>#4. #1 AND #2 AND #3</p> <p>#3. colorectal OR 'rectum'/exp OR rectum OR 'rectal'/exp OR rectal AND (neoplasm* OR 'cancer'/exp OR cancer OR 'cancers'/exp OR cancers OR carcinoma* OR carcinog* OR 'tumor'/exp OR tumor OR tumoral OR tumour* OR metastas* OR metastat* OR malign*)</p> <p>#2. 'rectum tumor'/exp AND [embase]/lim</p> <p>#1. 'colorectal tumor'/exp AND [embase]/lim</p>
Note	

Date	15 Jul 2010
Database (name + access; e.g.: Medline OVID)	Embase
Search Strategy (attention, for PubMed, check «Details»)	<p>#10. #4 AND #9</p> <p>#9. #5 AND prognostic AND scor* OR prognostic AND index OR 'nomogram' OR 'nomogram'/exp OR nomogram OR predictive AND ('model' OR 'model'/exp OR model) OR validation OR validate OR prognostic AND ('model' OR 'model'/exp OR model) OR predictor AND (score* OR scori* OR index OR 'model' OR 'model'/exp OR model OR predict* OR 'nomogram' OR 'nomogram'/exp OR nomogram OR validat*)</p> <p>#4. #1 AND #2 AND #3</p> <p>#3. colorectal OR 'rectum'/exp OR rectum OR 'rectal'/exp OR rectal AND (neoplasm* OR 'cancer'/exp OR cancer OR 'cancers'/exp OR cancers OR carcinoma* OR carcinog* OR 'tumor'/exp OR tumor OR tumoral OR tumour* OR metastas* OR metastat* OR malign*)</p> <p>#2. 'rectum tumor'/exp AND [embase]/lim</p> <p>#1. 'colorectal tumor'/exp AND [embase]/lim</p>
Note	

Date	July 27th, 2010
Database (name + access; e.g.: Medline OVID)	Cochrane Central Register of Controlled Trials
Search Strategy (attention, for PubMed, check «Details»)	<p>((colorectal OR rectum or rectal) AND (neoplas* OR cancer or cancers OR carcinom* OR carcinog* OR tumor OR tumors OR tumoral OR tumour* OR metastas* OR metastat* OR malign*)) AND (prognostic scor* OR prognostic index OR nomogram OR predictive model OR validation OR validate</p>

	OR prognostic model OR predictor OR prognosticator) AND (score* OR scori* OR index OR model OR predict* OR nomogram OR validat*) AND multivar*
Note	

2.2 PROGNOSTIC FACTORS

Survival	
Significant prognostic factors	Non-significant prognostic factors
<i>Clinical factors</i> <u>Patient</u> Bacterial translocation to lymph nodes BMI Health and physical subscale of QLI Insurance status Marital status Poor general condition/ Co-morbidity Socioeconomic status Venous tromboembolism <u>Tumor</u> Bowel obstruction Focal perforation Circulating Tumor Cells Local peritoneal involvement Lymphangitis carcinomatosa Serosal invasion Surgical curability Tumor regressing grading <u>Blood</u> Erythrocyte sedimentaion rate In vitro IL-6 production by peripheral blood mononuclear cell (PBMC) Natural Killer (NK)- cells Serum D-dimer Serum ferritin level Serum laminin <i>Pathological, genetic and molecular factors</i> A78-G/A7 reactivity Aberrant p16 methylation ABH isoantigens expression bcl2-reactivity CA IX expression CA72-4 expression CD8 expression CD31 expression	<i>Clinical factors</i> <u>Patient</u> Alcohol use Deprivation Ethnicity Global Quality of Life Peri-operative transfusion Personal history of cancer Sex <u>Tumor</u> Duration of symptoms Macroscopic aspect Size Blood Alkaline phosphatase Anemia Aspartate aminotransferase Erythrocyte sedimentation rate Serum IL-2 Serum IL-6 Serum TPA <i>Pathological, genetic and molecular factors</i> Adenomatous Polyposis Coli (APC)- mutation Cathepsin B level Fibrosis GST- α GST- μ Interleucin 10 (IL 10) expression Leptin expression LOH of 18q Loss of CDX2 expression Lymphocytic reaction Lymph vessel density Mucin 1 cell surface associated (MUC 1) expression Nuclear polarity

<p> CD34 expression Chomosomal Instability c-myc expression Cyclin A expression DCC protein expression DNA polymerase alpha positive cell rate E cadherin expression EGFR-expression Elevated binding of transcriptional regulators of u-PAR FADD-like IL-1β-converting enzyme (FLICE) inhibitory protein expression Glasgow Pognostic Score (GPS) Glutathione S-transferase (GST)-π expression GST-activity HCG- expression Heparanase expression HIF-1α expression kip1 expression KL-6 expression Klintrup criteria Loss of Heterozygosity (LOH) of 3p3 LOH of G219511 LOH of D3S647 Membrane Catenin expression Methylated HPP1 serum DNA Methylated HLTF serum DNA Mitotic Centromere-Associated Kinesin (MCAK) expression Mortalin expression mRNA level Myeloid differentiation factor 88 p21-ras expression p27 expression Pdcd4 expression Peritoneal cytology Perineural invasion Potential tumor doubling time Preoperative serum VEGF PTEN expression Raf kinase inhibitor (RKIP) expression Soluble urokinase-type Plasminogen Activator (suPAR) concentration sTie-2 receptor expression STMN1 expression Tetranectin expression Tissue Inhibitor of Metalloproteinase (TIMP-1) Tissue polypeptide antigen expression Tissue RNA of matrix metalloproteinase-9 </p>	<p> Peritumoural infiltration of granulocytes and lymphocytes P -glycoprotein expression pRB Proliferating Cell Nuclear Antigen(PCNA)index Survivin expression T antigen positivity Tissue Plasminogen Activator (TPA) in tissue Tn antigen positivity Thrombospondin 1 (TSP 1) Tubule configuration Tumor depth into mesorectum </p>
---	--

Toll-like receptor 4 expression Trypsin positivity VEGF-A in serum VEGF in tissue	
Prognostic factors with controversial significance	
<p><i>Clinical factors</i></p> Age Complication/Anastomotic leak Family history of cancer Location Serum CEA Stage (Dukes, Jass and TNM) Tobacco use/ Smoking behaviour <p><i>Pathological, genetic and molecular factors</i></p> Apoptotic index CA 19.9 in tissue CD8+/buds index CD44 expression Cyclin D overexpression Depth of invasion Differentiation/ Grade/ Growth pattern Histological type Ki-67 expression K-ras mutation Lymphatic infiltration Microsatellite Instability (MSI) status Microvascular Density / Tumor angiogenesis Nuclear staining density β -catenin p53 mutation Plasma VEGF-C level Platelet derived endothelial cells growth factor (PD-ECGF) Ploidy / DNA index Serum CA242 Sialyl Lex expression (SLX) Sialyl Tn immunoreactivity S-phase fraction labelling index / Duration of S-phase urokinase-type Plasminogen Activator receptor Vascular invasion White cell count/Neutrophils	
<i>Not patient specific</i>	
Chemotherapy way of administration Complexity of surgery Distal margin <1cm Mesorectal grade No of lymph nodes examined Surgical technique	Adjuvant therapy Chemotherapy Pathological circumferential resection margin (CRM) Type of first treatment

Treatment history	
Type of resection	
<i>Controversial not patient specific factors</i>	
Radiotherapy	

Local recurrence	
Significant prognostic factors	Non-significant prognostic factors
<i>Clinical factors</i> Gender Liver metastasis T-stage Distance from the anal verge <i>Pathological, genetic and molecular factors</i> Bcl-2 expression Microvessel density p53 (nuclear accumulation) PIK3CA mutation Ploidy S-phase fraction VEGF-C expression	<i>Pathological, genetic and molecular factors</i> Lymphatic involvement P glycoprotein expression
<i>Not patient specific</i>	
No Radiotherapy Perioperative blood transfusion	
Prognostic factors with controversial significance	
N-stage	

3 RESULTS

The primary search identified 981 articles: 926 in PubMed, 54 in Embase and 1 in the Cochrane Central Register of Controlled Trials. From this list, 308 articles were selected for full-text evaluation: 291 from PubMed, 16 from Embase and 1 from the Cochrane Central Register of Controlled Trials. After full-text evaluation, 152 articles were included in the final assessment. Reasons for exclusion are detailed in table 2.

Table 2: reason for exclusion

References excluded based on abstract and title		References excluded based on full text	
Reason	N	Reason	N
Population	355	Population	55
Outcome	242	Outcome	13
Design	74	Design	17
Other	0	Other	2
Total	672	Total	87

Details concerning the 152 retrieved papers are summarized in Table 3 (separate Excel file), and the global results from multivariate analyses are presented in Section 2.2. The main prognostic factor for overall survival is clearly related to the stage at presentation: patients with bowel obstruction, perforation, serosal invasion, or peritoneal metastasis fare worse. Gender does not seem to represent an independent prognostic factor, while the prognostic significance of age is variable among studies. Several studies have shown that socioeconomic deprivation represents an adverse prognostic factor for colorectal cancer survival. A wide array of pathological prognostic variables, macroscopic as well as microscopic and molecular, was identified. A number of recent studies has identified hospital volume as a prognostic factor in rectal cancer (Anwar 2010, Nugent 2010, Kressner 1998, Borowski 2010, van Gijn 2010).

Clinical and demographic variables with a impact on local recurrence include T stage, presence of liver metastasis, and gender. The impact of tumor location within the rectum on the risk of local recurrence is unclear at present, since some authors found a higher risk of local recurrence with low lying tumors (Faerden 2005) while others reported the opposite (Kusters 2009). Treatment-related factors influencing the risk of local recurrence include preoperative (chemo)radiation, performance of a total mesorectal excision (Pinsk 2007), and performance of abdominoperineal resection (den Dulk 2009). Among the pathological factors that may impact on local recurrence, the circumferential resection margin is clearly prominent (Bernstein 2009,

Quirke 2009). Finally, anastomotic leakage was shown in some reports to be associated with a higher risk of local recurrence (Eberhardt 2009, Law 2007). Several other reports, however, concluded that anastomotic leaks have no impact on local recurrence rate (Jörgren 2009, Bertelsen 2009, Lee 2008, Eriksen 2005).

There is very scarce literature on separate Quality of care indicators (QCI) previously identified in the setting of ProCare other than survival or local recurrence. Some specific factors are reported separately in appendix 2. The final report will tabulate relevant confounding factors for each QCI based on published evidence and on expert opinion from the participating clinicians.

The search including 'instrumental variable' as a term did not yield any results.

4 DISCUSSION

Several limitations apply to the interpretation of the present systematic literature search. First, most papers concern small patient numbers treated with a myriad of different therapeutic approaches and include colon as well as rectal cancer patients. The number of rectal cancer patients is usually not specified or a (small) minority of the overall population. This is relevant since the biological behavior of (low) rectal cancer and the paramount importance of surgical technique in achieving the desired outcome are quite different compared to colon cancer. As there are only 23 studies on rectal cancer alone, studies on colorectal cancer were nevertheless included. Second, almost all data were the result of retrospective studies. Studies not including some form of multivariate analysis were excluded. This criterion was maintained in order to guarantee a minimal quality of included studies.

It is important to note that most papers study prognostic factors through joint regression models, which contain the patient-specific variables available. Whether a particular variable enters as a significant predictor into such joint model will greatly depend on which other variables are further included in the model. Indeed, both the magnitude and even the sign of the true effect on outcome may change depending on which other factors are entered. For some sets of variables only one may need to be appropriately corrected for the prognostic value involved, i.e. they can act as each other's surrogate in this sense. This could imply that as soon as one is entered, the other variables no longer have anything to add. Which of them actually enters may then be a matter of chance. This complicates the definition and role of the prognostic factors for reporting purposes. Beyond the magnitude of its systematic effect in the joint model, there is also the issue of precision. Whether a particular factor (in a joint or univariate model) is significant or not, not only depends on the magnitude of its systematic effect, but also on the precision with which it is estimated and hence on the sample size and covariate distribution in the studied population. In the light of this, and the fact that current and future sets of available covariates may rarely overlap exactly with what is reported in the literature, we will report here first on any variable found to be a significant prognostic factor. In the more detailed report we will indicate in what combination of covariates it occurred with what weight.

5 KEY POINTS

- The primary search identified 981 articles. From this list, 308 articles were selected for full-text evaluation leading to 152 articles included in the final assessment. From these articles, an extensive list of prognostic factors for overall survival was obtained as well as a less extensive list of prognostic factors for local recurrence, cancer-specific survival and post-operative complications. There is very scarce literature on prognostic factors for other QCIIs identified in the setting of PROCARE.
- The literature search imposed restrictions in terms of study design and patient population. Since a mere 23 studies considered just rectal cancer patients, also studies on colon cancer patients were eligible for our selection.
- Most papers study prognostic factors through multivariate regression models, hence the direction and magnitude of effect of a specific prognostic factor on the outcome depends heavily on the other factors included in the model.

Cancer specific survival	
Significant prognostic factors	Non-significant prognostic factors
<i>Clinical factors</i> Age BMI Recurrence Stage <i>Pathological, genetic and molecular factors</i> CD44v6 Differentiation Glasgow Pognostic Score (GPS) Klintrup criteria Pattern of growth Tumor budding Tumor infiltrating lymphocytes urokinase-type Plasminogen Activator urokinase-type Plasminogen Activator receptor	<i>Pathological and molecular factors</i> MMP-9 Pattern of differentiation

Postoperative complications	
Significant prognostic factors	Non-significant prognostic factors
<i>Clinical factors</i> Age Malnutrition <i>Pathological, genetic and molecular factors</i> SF-36 (social functioning)	<i>Clinical factors</i> AJCC stage ASA class Obesity Race Residence
<i>Not patient specific</i>	
Intraoperative contamination	Centre case volume Operative technique

REFERENCES

- Abubaker, J., P. Bavi, et al. (2009). "Prognostic significance of alterations in KRAS isoforms KRAS-4A/4B and KRAS mutations in colorectal carcinoma." *J Pathol* 219(4): 435-45.
- Acikalin, M. F., U. Oner, et al. (2005). "Tumour angiogenesis and mast cell density in the prognostic assessment of colorectal carcinomas." *Dig Liver Dis* 37(3): 162-9
- Adam, I. J. and M. O. Mohamdee (1994). "Role of circumferential margin involvement in the local recurrence of rectal cancer." *Lancet* 344(8924): 707.
- Al-Mulla, F., S. Hagan, et al. (2006). "Raf kinase inhibitor protein expression in a survival analysis of colorectal cancer patients." *J Clin Oncol* 24(36): 5672-9.
- Alcalay, A., T. Wun, et al. (2006). "Venous thromboembolism in patients with colorectal cancer: incidence and effect on survival." *J Clin Oncol* 24(7): 1112-8.
- Amato, A. and M. Pescatori (2006) Perioperative blood transfusions for the recurrence of colorectal cancer. *Cochrane Database of Systematic Reviews* DOI: 10.1002/14651858.CD005033.pub2
- Anwar S, Fraser S, Hill J. Surgical specialization and training - its relation to clinical outcome for colorectal cancer surgery. *J Eval Clin Pract*. 2010 Aug 4
- Anthony, T., L. S. Hynan, et al. (2003). "The association of pretreatment health-related quality of life with surgical complications for patients undergoing open surgical resection for colorectal cancer." *Ann Surg* 238(5): 690-6.
- Armitage, N. C., K. C. Ballantyne, et al. (1990). "The influence of tumour cell DNA content on survival in colorectal cancer: a detailed analysis." *Br J Cancer* 62(5): 852-6.

- Asteria, C. R., G. Gagliardi, et al. (2008). "Anastomotic leaks after anterior resection for mid and low rectal cancer: survey of the Italian Society of Colorectal Surgery." *Tech Coloproctol* 12(2): 103-10.
- Asghari-Jafarabadi, M., E. Hajizadeh, et al. (2009). "Site-specific evaluation of prognostic factors on survival in Iranian colorectal cancer patients: a competing risks survival analysis." *Asian Pac J Cancer Prev* 10(5): 815-21.
- Bendardaf, R., A. Buhmeida, et al. (2010) "MMP-9 (gelatinase B) expression is associated with disease-free survival and disease-specific survival in colorectal cancer patients." *Cancer Invest* 28(1): 38-43.
- Benatti, P., R. Gafa, et al. (2005). "Microsatellite instability and colorectal cancer prognosis." *Clin Cancer Res* 11(23): 8332-40.
- Bell, S. W., K. G. Walker, et al. (2003). "Anastomotic leakage after curative anterior resection results in a higher prevalence of local recurrence." *Br J Surg* 90(10): 1261-6
- Baskaranathan, S., J. Philips, et al. (2004). "Free colorectal cancer cells on the peritoneal surface: correlation with pathologic variables and survival." *Dis Colon Rectum* 47(12): 2076-9.
- Bardi, G., C. Fenger, et al. (2004). "Tumor karyotype predicts clinical outcome in colorectal cancer patients." *J Clin Oncol* 22(13): 2623-34.
- Banerjea, A., R. E. Hands, et al. (2009). "Microsatellite and chromosomal stable colorectal cancers demonstrate poor immunogenicity and early disease recurrence." *Colorectal Dis* 11(6): 601-8.
- Baldus, S. E., T. K. Zirbes, et al. (2000). "Thomsen-Friedenreich antigen presents as a prognostic factor in colorectal carcinoma: A clinicopathologic study of 264 patients." *Cancer* 88(7): 1536-43.
- Bahnassy, A. A., A. R. Zekri, et al. (2004). "Cyclin A and cyclin D1 as significant prognostic markers in colorectal cancer patients." *BMC Gastroenterol* 4: 22.
- Baba, Y., K. Nosho, et al. (2009). "Relationship of CDX2 loss with molecular features and prognosis in colorectal cancer." *Clin Cancer Res* 15(14): 4665-73.
- Belluco, C., G. Esposito, et al. (1999). "Absence of the cell cycle inhibitor p27(Kip1) protein predicts poor outcome in patients with stage I-III colorectal cancer." *Annals of Surgical Oncology* 6(1): 19-25.

- Belluco, C., D. Nitti, et al. (2000). "Interleukin-6 blood level is associated with circulating carcinoembryonic antigen and prognosis in patients with colorectal cancer." *Annals of Surgical Oncology* 7(2): 133-138.
- Bernstein et al.(2009) [Circumferential resection margin as a prognostic factor in rectal cancer](#) *Br J Surg.* Nov;96(11):1348-57
- Bertelsen CA, Andreassen AH, Jørgensen T, Harling H; on behalf of the Danish Colorectal Cancer Group. Anastomotic leakage after curative anterior resection for rectal cancer: short and long term outcome. *Colorectal Dis.* 2009 Apr 29
- Bondi, J., M. Pretorius, et al. (2009). "Large-scale genomic instability in colon adenocarcinomas and correlation with patient outcome." *APMIS* 117(10): 730-6.
- Bosch, B., U. Guller, et al. (2003). "Perioperative detection of disseminated tumour cells is an independent prognostic factor in patients with colorectal cancer." *Br J Surg* 90(7): 882-8.
- Broll, R., P. Busch, et al. (2005). "Influence of thymidylate synthase and p53 protein expression on clinical outcome in patients with colorectal cancer." *Int J Colorectal Dis* 20(2): 94-102.
- Borowski et al., (2010) Volume-outcome analysis of colorectal cancer-related outcomes *British Journal of Surgery* 97: 1416–1430
- Cianchi, F., A. Palomba, et al. (2002). "Lymph node recovery from colorectal tumor specimens: Recommendation for a minimum number of lymph nodes to be examined." *World Journal of Surgery* 26(3): 384-389.
- Cao, D., M. Hou, et al. (2009). "Expression of HIF-1alpha and VEGF in colorectal cancer: association with clinical outcomes and prognostic implications." *BMC Cancer* 9: 432.
- Carpelan-Holmstrom, M., C. Haglund, et al. (1996). "Independent prognostic value of preoperative serum markers CA 242, specific tissue polypeptide antigen and human chorionic gonadotrophin beta, but not of carcinoembryonic antigen or tissue polypeptide antigen in colorectal cancer." *Br J Cancer* 74(6): 925-9.
- Chang, S. C., J. K. Lin, et al. (2006). "Relationship between genetic alterations and prognosis in sporadic colorectal cancer." *Int J Cancer* 118(7): 1721-7.
- Cheah, P. Y., P. H. Choo, et al. (2002). "A survival-stratification model of human colorectal carcinomas with beta-catenin and p27kip1." *Cancer* 95(12): 2479-86.

- Chin, K. F., J. Greenman, et al. (2000). "Pre-operative serum vascular endothelial growth factor can select patients for adjuvant treatment after curative resection in colorectal cancer." *Br J Cancer* 83(11): 1425-31.
- Chin, K. F., J. Greenman, et al. (2003). "Vascular endothelial growth factor and soluble Tie-2 receptor in colorectal cancer: associations with disease recurrence." *Eur J Surg Oncol* 29(6): 497-505.
- Chin, K. F., R. Kallam, et al. (2007). "Bacterial translocation may influence the long-term survival in colorectal cancer patients." *Dis Colon Rectum* 50(3): 323-30.
- Choi, H. J., M. S. Hyun, et al. (1998). "Tumor angiogenesis as a prognostic predictor in colorectal carcinoma with special reference to mode of metastasis and recurrence." *Oncology* 55(6): 575-81.
- Choi, S. W., K. J. Lee, et al. (2002). "Genetic classification of colorectal cancer based on chromosomal loss and microsatellite instability predicts survival." *Clin Cancer Res* 8(7): 2311-22.
- Clinchy, B., A. Fransson, et al. (2007). "Preoperative interleukin-6 production by mononuclear blood cells predicts survival after radical surgery for colorectal carcinoma." *Cancer* 109(9): 1742-9.
- Colakoglu, T., S. Yildirim, et al. (2008). "Clinicopathological significance of PTEN loss and the phosphoinositide 3-kinase/Akt pathway in sporadic colorectal neoplasms: is PTEN loss predictor of local recurrence?" *Am J Surg* 195(6): 719-25.
- Conlin, A., G. Smith, et al. (2005). "The prognostic significance of K-ras, p53, and APC mutations in colorectal carcinoma." *Gut* 54(9): 1283-6.
- den Dulk et al., (2009) [The abdominoperineal resection itself is associated with an adverse outcome: The European experience based on a pooled analysis of five European randomised clinical trials on rectal cancer](#) *Eur J Cancer* 45: 1175
- Dundas, S. R., L. C. Lawrie, et al. (2005). "Mortalin is over-expressed by colorectal adenocarcinomas and correlates with poor survival." *Journal of Pathology* 205(1): 74-81.
- Dimitriadis, E., T. Trangas, et al. (2007). "Expression of oncofetal RNA-binding protein CRD-BP/IMP1 predicts clinical outcome in colon cancer." *Int J Cancer* 121(3): 486-94.
- Duncan, T. J., N. F. Watson, et al. (2007). "The role of MUC1 and MUC3 in the biology and prognosis of colorectal cancer." *World J Surg Oncol* 5: 31.

- Eberhardt JM et al.,(2009) [The Impact of Anastomotic Leak and Intra-Abdominal Abscess on Cancer-Related Outcomes After Resection for Colorectal Cancer: A Case Control Study](#) Dis Colon Rectum. Mar;52(3):380-6;
- Elzagheid, A., A. Algars, et al. (2006). "E-cadherin expression pattern in primary colorectal carcinomas and their metastases reflects disease outcome." World J Gastroenterol 12(27): 4304-9.
- Engel, C. J., S. T. Bennett, et al. (1996). "Tumor angiogenesis predicts recurrence in invasive colorectal cancer when controlled for Dukes staging." Am J Surg Pathol 20(10): 1260-5.
- Eriksen MT, Wibe A, Norstein J, Haffner J, Wiig JN; Norwegian Rectal Cancer Group. Anastomotic leakage following routine mesorectal excision for rectal cancer in a national cohort of patients. Colorectal Dis. 2005 Jan;7(1):51-7
- Filiz, A. I., I. Sucullu, et al. (2009). "Persistent high postoperative carcinoembryonic antigen in colorectal cancer patients--is it important?" Clinics (Sao Paulo) 64(4): 287-94.
- Faerden et al. (2005) Title: [Total mesorectal excision for rectal cancer: Difference in outcome for low and high rectal cancer](#) Dis Colon Rectum 48: 2224
- Herbst, A., M. Wallner, et al. (2009). "Methylation of helicase-like transcription factor in serum of patients with colorectal cancer is an independent predictor of disease recurrence." Eur J Gastroenterol Hepatol 21(5): 565-9.
- Hermanek, P., Jr., I. Guggenmoos-Holzmanna, et al. (1989). "[Effect of the transfusion of blood and hemoderivatives on the prognosis of colorectal cancer]." Langenbecks Arch Chir 374(2): 118-24.
- Hermanek, P., W. Gonselmann, et al. (1981). "[Prediction of local recurrences after surgery of carcinoma of the middle rectum (author's transl)]." Langenbecks Arch Chir 354(2): 133-46.
- Hiraga, Y., S. Tanaka, et al. (1998). "Immunoreactive MUC1 expression at the deepest invasive portion correlates with prognosis of colorectal cancer." Oncology 55(4): 307-19.
- Hogdall, C. K., I. J. Christensen, et al. (2002). "Serum tetranectin is an independent prognostic marker in colorectal cancer and weakly correlated with plasma suPAR, plasma PAI-1 and serum CEA." APMIS 110(9): 630-8.

- Holten-Andersen, M., I. J. Christensen, et al. (2004). "Association between preoperative plasma levels of tissue inhibitor of metalloproteinases 1 and rectal cancer patient survival. a validation study." *Eur J Cancer* 40(1): 64-72.
- Huguier, M., F. Depoux, et al. (1990). "Adenocarcinoma of the rectum treated by abdominoperineal excision: multivariate analysis of prognostic factors." *Int J Colorectal Dis* 5(3): 144-7.
- Huh, J. W., H. R. Kim, et al. (2009). "Expression of standard CD44 in human colorectal carcinoma: association with prognosis." *Pathol Int* 59(4): 241-6.
- Ihmann, T., J. Liu, et al. (2004). "High-level mRNA quantification of proliferation marker pKi-67 is correlated with favorable prognosis in colorectal carcinoma." *J Cancer Res Clin Oncol* 130(12): 749-56.
- Iniesta, P., M. J. Massa, et al. (2000). "Loss of heterozygosity at 3p23 is correlated with poor survival in patients with colorectal carcinoma." *Cancer* 89(6): 1220-7.
- Ishikawa, H., H. Fujii, et al. (1999). "Tumor angiogenesis predicts recurrence with normal serum carcinoembryonic antigen in advanced rectal carcinoma patients." *Surg Today* 29(10): 983-91.
- Ishikawa, K., Y. Kamohara, et al. (2008). "Mitotic centromere-associated kinesin is a novel marker for prognosis and lymph node metastasis in colorectal cancer." *Br J Cancer* 98(11): 1824-9.
- Itzkowitz, S. H., E. J. Bloom, et al. (1990). "Sialosyl-Tn. A novel mucin antigen associated with prognosis in colorectal cancer patients." *Cancer* 66(9): 1960-6.
- Jass, J. R. (1986). "Lymphocytic infiltration and survival in rectal cancer." *J Clin Pathol* 39(6): 585-9.
- Jatzko, G. R., M. Jagoditsch, et al. (1999). "Long-term results of radical surgery for rectal cancer: multivariate analysis of prognostic factors influencing survival and local recurrence." *Eur J Surg Oncol* 25(3): 284-91.
- Jörgren F, Johansson R, Damber L, Lindmark G. Risk Factors of Rectal Cancer. Local Recurrence: Population-based Survey and Validation of the Swedish Rectal Cancer Registry. *Colorectal Dis*. 2009 Apr 27
- Kahlenberg, M. S., D. L. Stoler, et al. (2000). "p53 tumor suppressor gene mutations predict decreased survival of patients with sporadic colorectal carcinoma." *Cancer* 88(8): 1814-9.

- Kaio, E., S. Tanaka, et al. (2003). "Clinical significance of angiogenic factor expression at the deepest invasive site of advanced colorectal carcinoma." *Oncology* 64(1): 61-73.
- Kanazawa, H., H. Mitomi, et al. (2008). "Tumour budding at invasive margins and outcome in colorectal cancer." *Colorectal Dis* 10(1): 41-7.
- Kawasaki, H., D. C. Altieri, et al. (1998). "Inhibition of apoptosis by survivin predicts shorter survival rates in colorectal cancer." *Cancer Res* 58(22): 5071-4.
- Kido, A., M. Mori, et al. (1996). "Immunohistochemical expression of beta-human chorionic gonadotropin in colorectal carcinoma." *Surg Today* 26(12): 966-70.
- Kopp, R., E. Rothbauer, et al. (2003). "Reduced survival of rectal cancer patients with increased tumor epidermal growth factor receptor levels." *Dis Colon Rectum* 46(10): 1391-9.
- Korkeila, E., K. Talvinen, et al. (2009). "Expression of carbonic anhydrase IX suggests poor outcome in rectal cancer." *Br J Cancer* 100(6): 874-80.
- Kos, J., H. J. Nielsen, et al. (1998). "Prognostic values of cathepsin B and carcinoembryonic antigen in sera of patients with colorectal cancer." *Clin Cancer Res* 4(6): 1511-6.
- Kressner, U., B. Glimelius, et al. (1998). "Increased serum p53 antibody levels indicate poor prognosis in patients with colorectal cancer." *Br J Cancer* 77(11): 1848-51.
- Kokal, W., K. Sheibani, et al. (1986). "TUMOR DNA CONTENT IN THE PROGNOSIS OF COLORECTAL-CARCINOMA." *Jama-Journal of the American Medical Association* 255(22): 3123-3127.
- Kusters et al. (2009) [Patterns of local recurrence in locally advanced rectal cancer after intra-operative radiotherapy containing multimodality treatment](#) *Radiother Oncol* 92: 221
- Law et al. (2007) : [Anastomotic leakage is associated with poor long-term outcome in patients after curative colorectal resection for malignancy](#) *J Gastrointest Surg.* Jan;11(1):8-15
- Le Voyer, T. E., E. R. Sigurdson, et al. (2003). "Colon cancer survival is associated with increasing number of lymph nodes analyzed: A secondary survey of Intergroup trial INT-0089." *Journal of Clinical Oncology* 21(15): 2912-2919.

- Lamberti, C., S. Lundin, et al. (2007). "Microsatellite instability did not predict individual survival of unselected patients with colorectal cancer." *Int J Colorectal Dis* 22(2): 145-52.
- Lee WS, Yun SH, Roh YN, Yun HR, Lee WY, Cho YB, Chun HK. Risk factors and clinical outcome for anastomotic leakage after total mesorectal excision for rectal cancer. *World J Surg*. 2008 Jun;32(6):1124-9
- Li, M., J. Y. Li, et al. (2009). "Comparison of carcinoembryonic antigen prognostic value in serum and tumour tissue of patients with colorectal cancer." *Colorectal Dis* 11(3): 276-81.
- Liebig, C., G. Ayala, et al. (2009). "Perineural invasion is an independent predictor of outcome in colorectal cancer." *J Clin Oncol* 27(31): 5131-7.
- Lin, J. K., S. C. Chang, et al. (2003). "Prognostic value of DNA ploidy patterns of colorectal adenocarcinoma." *Hepatogastroenterology* 50(54): 1927-32.
- Lin, M., S. P. Ma, et al. "Intratumoral as well as peritumoral lymphatic vessel invasion correlates with lymph node metastasis and unfavourable outcome in colorectal cancer." *Clin Exp Metastasis* 27(3): 123-32.
- Lindmark, G., B. Gerdin, et al. (1994). "Prognostic predictors in colorectal cancer." *Dis Colon Rectum* 37(12): 1219-27.
- Lis, C. G., D. Gupta, et al. (2006). "Can patient satisfaction with quality of life predict survival in advanced colorectal cancer?" *Support Care Cancer* 14(11): 1104-10.
- Lorenzi, M., B. Lorenzi, et al. (2006). "Serum ferritin in colorectal cancer patients and its prognostic evaluation." *Int J Biol Markers* 21(4): 235-41.
- Louhimo, J., M. Carpelan-Holmstrom, et al. (2002). "Serum HCG beta, CA 72-4 and CEA are independent prognostic factors in colorectal cancer." *Int J Cancer* 101(6): 545-8.
- Lugli, A., E. Karamitopoulou, et al. (2009). "CD8+ lymphocytes/ tumour-budding index: an independent prognostic factor representing a 'pro-/anti-tumour' approach to tumour host interaction in colorectal cancer." *Br J Cancer* 101(8): 1382-92.
- Lundin, M., S. Nordling, et al. (1999). "Sialyl Tn is a frequently expressed antigen in colorectal cancer: No correlation with patient prognosis." *Oncology* 57(1): 70-6.
- Maeda, K., Y. Chung, et al. (1998). "Cyclin D1 overexpression and prognosis in colorectal adenocarcinoma." *Oncology* 55(2): 145-51.

- Maslekar, S., A. Sharma, et al. (2007). "Mesorectal grades predict recurrences after curative resection for rectal cancer." *Dis Colon Rectum* 50(2): 168-75.
- Maurer, G. D., J. H. Leupold, et al. (2007). "Analysis of specific transcriptional regulators as early predictors of independent prognostic relevance in resected colorectal cancer." *Clin Cancer Res* 13(4): 1123-32.
- Michel, P., M. Paresy, et al. (2000). "Pre-operative kinetic parameter determination of colorectal adenocarcinomas. Prognostic significance." *Eur J Gastroenterol Hepatol* 12(3): 275-80.
- Mitomi, H., N. Fukui, et al. "Aberrant p16((INK4a)) methylation is a frequent event in colorectal cancers: prognostic value and relation to mRNA expression and immunoreactivity." *J Cancer Res Clin Oncol* 136(2): 323-31.
- Miyazaki, T., N. Okada, et al. (2008). "Clinical significance of plasma level of vascular endothelial growth factor-C in patients with colorectal cancer." *Jpn J Clin Oncol* 38(12): 839-43.
- Moghimidehkhordi, B., A. Safaee, et al. (2008). "Prognostic factors in 1,138 Iranian colorectal cancer patients." *Int J Colorectal Dis* 23(7): 683-8.
- Molland, G., O. F. Dent, et al. (1995). "Transfusion does not influence patient survival after resection of colorectal cancer." *Aust N Z J Surg* 65(8): 592-5.
- Monnet, E., J. Faivre, et al. (1999). "Influence of stage at diagnosis on survival differences for rectal cancer in three European populations." *Br J Cancer* 81(3): 463-8.
- Mudduluru, G., F. Medved, et al. (2007). "Loss of programmed cell death 4 expression marks adenoma-carcinoma transition, correlates inversely with phosphorylated protein kinase B, and is an independent prognostic factor in resected colorectal cancer." *Cancer* 110(8): 1697-707.
- Mulder, J. W., I. O. Baas, et al. (1995). "Evaluation of p53 protein expression as a marker for long-term prognosis in colorectal carcinoma." *Br J Cancer* 71(6): 1257-62.
- Mulder, T. P., H. W. Verspaget, et al. (1995). "Glutathione S-transferase pi in colorectal tumors is predictive for overall survival." *Cancer Res* 55(12): 2696-702.
- Munro, A. J., A. H. Bentley, et al. (2006). "Smoking compromises cause-specific survival in patients with operable colorectal cancer." *Clin Oncol (R Coll Radiol)* 18(6): 436-40.

- Nakagoe, T., K. Fukushima, et al. (1993). "Immunohistochemical expression of sialyl Lex antigen in relation to survival of patients with colorectal carcinoma." *Cancer* 72(8): 2323-30.
- Nakagoe, T., A. Nanashima, et al. (2000). "Expression of blood group antigens A, B and H in carcinoma tissue correlates with a poor prognosis for colorectal cancer patients." *J Cancer Res Clin Oncol* 126(7): 375-82.
- Nakagoe, T., T. Sawai, et al. (2000). "Prognostic value of circulating sialyl Tn antigen in colorectal cancer patients." *Anticancer Res* 20(5C): 3863-9.
- Nakayama, T., M. Watanabe, et al. (1997). "CA19-9 as a predictor of recurrence in patients with colorectal cancer." *J Surg Oncol* 66(4): 238-43.
- Neoptolemos, J. P., G. D. Oates, et al. (1995). "Cyclin/proliferation cell nuclear antigen immunohistochemistry does not improve the prognostic power of Dukes' or Jass' classifications for colorectal cancer." *Br J Surg* 82(2): 184-7.
- Nespoli, A., L. Gianotti, et al. (2006). "Impact of postoperative infections on survival in colon cancer patients." *Surg Infect (Larchmt)* 7 Suppl 2: S41-3.
- Nobuhisa, T., Y. Naomoto, et al. (2005). "Heparanase expression correlates with malignant potential in human colon cancer." *J Cancer Res Clin Oncol* 131(4): 229-37.
- Noda, E., K. Maeda, et al. (2007). "Predictive value of vascular endothelial growth factor-C expression for local recurrence of rectal carcinoma." *Oncol Rep* 17(6): 1327-31.
- Noguchi, T., R. Kikuchi, et al. (2003). "Prognostic significance of p27/kip1 and apoptosis in patients with colorectal carcinoma." *Oncol Rep* 10(4): 827-31.
- Noura, S., M. Ohue, et al. (2009). "Long-term prognostic value of conventional peritoneal lavage cytology in patients undergoing curative colorectal cancer resection." *Dis Colon Rectum* 52(7): 1312-20.
- Nagtegaal, I. D., C. A. A. Marijnen, et al. (2002). "Circumferential margin involvement is still an important predictor of local recurrence in rectal carcinoma - Not one millimeter but two millimeters is the limit." *American Journal of Surgical Pathology* 26(3): 350-357
- Nagtegaal, I. D. and P. Quirke (2007). "Colorectal tumour deposits in the mesorectum and pericolon; a critical review." *Histopathology* 51(2): 141-149

- Nakagoe, T., T. Sawai, et al. (2000). "Difference in prognostic value between sialyl Lewis^x and sialyl Lewis(x) antigens in blood samples obtained from the drainage veins of the colorectal tumors." *Letters* 159(2): 159-168.
- Nugent E, Neary P. Rectal cancer surgery: volume-outcome analysis. *Int J Colorectal Dis*. 2010 Jul 27. [Epub ahead of print]
- Ofner, D., K. Riehemann, et al. (1995). "Immunohistochemically detectable bcl-2 expression in colorectal carcinoma: correlation with tumour stage and patient survival." *Br J Cancer* 72(4): 981-5.
- Ogino, S., K. Nosho, et al. (2009). "A cohort study of STMN1 expression in colorectal cancer: body mass index and prognosis." *Am J Gastroenterol* 104(8): 2047-56.
- Ogino, S., K. Nosho, et al. (2009). "Prognostic significance and molecular associations of 18q loss of heterozygosity: a cohort study of microsatellite stable colorectal cancers." *J Clin Oncol* 27(27): 4591-8.
- Oya, M., Y. Akiyama, et al. (2001). "High preoperative plasma D-dimer level is associated with advanced tumor stage and short survival after curative resection in patients with colorectal cancer." *Jpn J Clin Oncol* 31(8): 388-94.
- Paik, S. S., S. M. Jang, et al. (2009). "Leptin expression correlates with favorable clinicopathologic phenotype and better prognosis in colorectal adenocarcinoma." *Ann Surg Oncol* 16(2): 297-303.
- Palmqvist, R., R. Stenling, et al. (1999). "Prognostic significance of p27(Kip1) expression in colorectal cancer: a clinico-pathological characterization." *J Pathol* 188(1): 18-23.
- Park, Y. J., E. G. Youk, et al. (1999). "Experience of 1446 rectal cancer patients in Korea and analysis of prognostic factors." *Int J Colorectal Dis* 14(2): 101-6.
- Pinsk,(2007) [Total mesorectal excision and management of rectal cancer](#) Expert Rev Anticancer Ther10: 1395
- Quirke et al. *Lancet*. (2009) [Effect of the plane of surgery achieved on local recurrence in patients with operable rectal cancer: a prospective study using data from the MRC CR07 and NCIC-CTG CO16 randomised clinical trial](#) Mar 7;373(9666):821-8
- Ratto, C., R. Ricci, et al. (2002). "Mesorectal microfoci adversely affect the prognosis of patients with rectal cancer." *Diseases of the Colon & Rectum* 45(6): 733-742.

- Ruo, L. Y., S. Tickoo, et al. (2002). "Long-term prognostic significance of extent of rectal cancer response to preoperative radiation and chemotherapy." *Annals of Surgery* 236(1): 75-81.
- Robbins, A. S., A. L. Pavluck, et al. (2009). "Insurance status, comorbidity level, and survival among colorectal cancer patients age 18 to 64 years in the National Cancer Data Base from 2003 to 2005." *J Clin Oncol* 27(22): 3627-33.
- Roncucci, L., R. Fante, et al. (1996). "Survival for colon and rectal cancer in a population-based cancer registry." *Eur J Cancer* 32A(2): 295-302.
- Ropponen, K. M., M. J. Eskelinen, et al. (1998). "Expression of CD44 and variant proteins in human colorectal cancer and its relevance for prognosis." *Scand J Gastroenterol* 33(3): 301-9.
- Roxburgh, C. S., J. M. Salmond, et al. (2009). "Comparison of the prognostic value of inflammation-based pathologic and biochemical criteria in patients undergoing potentially curative resection for colorectal cancer." *Ann Surg* 249(5): 788-93.
- Saito, M., A. Yamaguchi, et al. (1999). "Expression of DCC protein in colorectal tumors and its relationship to tumor progression and metastasis." *Oncology* 56(2): 134-41.
- Salud, A., J. M. Porcel, et al. (1999). "Prognostic significance of DNA ploidy, S-phase fraction, and P-glycoprotein expression in colorectal cancer." *J Surg Oncol* 72(3): 167-74.
- Shepherd, N. A., K. J. Baxter, et al. (1997). "The prognostic importance of peritoneal involvement in colonic cancer: a prospective evaluation." *Gastroenterology* 112(4): 1096-102.
- Shibata, D., M. A. Reale, et al. (1996). "The DCC protein and prognosis in colorectal cancer." *New England Journal of Medicine* 335(23): 1727-1732.
- Stephens, R. W., H. J. Nielsen, et al. (1999). "Plasma urokinase receptor levels in patients with colorectal cancer: relationship to prognosis." *J Natl Cancer Inst* 91(10): 869-74.
- Szynglarewicz, B., R. Matkowski, et al. (2007). "Clinical factors in prediction of prognosis after anterior resection with total mesorectal excision for carcinoma of the rectum." *Oncol Rep* 17(2): 471-5.
- Szynglarewicz, B., R. Matkowski, et al. (2007). "Predictive value of lymphocytic infiltration and character of invasive margin following total mesorectal excision with

- sphincter preservation for the high-risk carcinoma of the rectum." *Adv Med Sci* 52: 159-63.
- Tanaka, M., K. Omura, et al. (1994). "Prognostic factors of colorectal cancer: K-ras mutation, overexpression of the p53 protein, and cell proliferative activity." *J Surg Oncol* 57(1): 57-64.
- Tanigawa, N., H. Amaya, et al. (1997). "Tumor angiogenesis and mode of metastas
- Tepper, J. E., M. J. O'Connell, et al. (2001). "Impact of number of nodes retrieved on outcome in patients with rectal cancer." *Journal of Clinical Oncology* 19(1): 157-163.
- Tomoda, H., H. Baba, et al. (1998). "DNA index as a significant predictor of recurrence in colorectal cancer." *Dis Colon Rectum* 41(3): 286-90.
- Tsuchiya, A., Y. Ando, et al. (1992). "Flow cytometric DNA analysis in Japanese colorectal cancer. A multivariate analysis." *Eur J Surg Oncol* 18(6): 585-90.
- Tsuji, T., T. Sawai, et al. (2004). "Platelet-derived endothelial cell growth factor expression is an independent prognostic factor in colorectal cancer patients after curative surgery." *Eur J Surg Oncol* 30(3): 296-302.
- Uen, Y. H., C. Y. Lu, et al. (2008). "Persistent presence of postoperative circulating tumor cells is a poor prognostic factor for patients with stage I-III colorectal cancer after curative resection." *Ann Surg Oncol* 15(8): 2120-8.
- Ullenhag, G. J., A. Mukherjee, et al. (2007). "Overexpression of FLIPL is an independent marker of poor prognosis in colorectal cancer patients." *Clin Cancer Res* 13(17): 5070-5.
- Valera, V., N. Yokoyama, et al. (2005). "Clinical significance of Ki-67 proliferation index in disease progression and prognosis of patients with resected colorectal carcinoma." *Br J Surg* 92(8): 1002-7.
- van Gijn et al., (2010) "Volume and outcome in colorectal cancer surgery" *Eur J Surg Oncol* 36: S55-S63
- Vecchio, F. M., V. Valentini, et al. (2005). "The relationship of pathologic tumor regression grade (TRG) and outcomes after preoperative therapy in rectal cancer." *Int J Radiat Oncol Biol Phys* 62(3): 752-60.
- Visca, P., P. L. Alo, et al. (1999). "Immunohistochemical expression of fatty acid synthase, apoptotic- regulating genes, proliferating factors, and ras protein

- product in colorectal adenomas, carcinomas, and adjacent nonneoplastic mucosa." *Clinical Research* 5(12): 4111-4118.
- Wallner, M., A. Herbst, et al. (2006). "Methylation of serum DNA is an independent prognostic marker in colorectal cancer." *Clin Cancer Res* 12(24): 7347-52.
- Wang, E. L., Z. R. Qian, et al. "High expression of Toll-like receptor 4/myeloid differentiation factor 88 signals correlates with poor prognosis in colorectal cancer." *Br J Cancer* 102(5): 908-15.
- Wang, W. S., J. K. Lin, et al. (2000). "Preoperative carcinoembryonic antigen level as an independent prognostic factor in colorectal cancer: Taiwan experience." *Jpn J Clin Oncol* 30(1): 12-6.
- Wanitsuwan, W., S. Kanngurn, et al. (2008). "Overall expression of beta-catenin outperforms its nuclear accumulation in predicting outcomes of colorectal cancers." *World J Gastroenterol* 14(39): 6052-9.
- Werther, K., I. J. Christensen, et al. (2002). "Prognostic impact of matched preoperative plasma and serum VEGF in patients with primary colorectal carcinoma." *Br J Cancer* 86(3): 417-23.
- West, N. P., M. Dattani, et al. "The proportion of tumour cells is an independent predictor for survival in colorectal cancer patients." *Br J Cancer* 102(10): 1519-23.
- Yamaguchi, A., T. Goi, et al. (1998). "Clinical significance of combined immunohistochemical detection of CD44v and sialyl LeX expression for colorectal cancer patients undergoing curative resection." *Oncology* 55(5): 400-3.
- Yamaguchi, A., Y. Hirono, et al. (1992). "DNA polymerase alpha positive-cell rate in colorectal cancer and its relationship to prognosis." *Br J Cancer* 65(3): 421-4.
- Yamaguchi, A., T. Urano, et al. (1996). "Expression of a CD44 variant containing exons 8 to 10 is a useful independent factor for the prediction of prognosis in colorectal cancer patients." *J Clin Oncol* 14(4): 1122-7.
- Yamamoto, H., S. Iku, et al. (2003). "Association of trypsin expression with tumour progression and matrilysin expression in human colorectal cancer." *J Pathol* 199(2): 176-84.
- Yamamoto, S., T. Akasu, et al. (2003). "Long-term prognostic value of conventional peritoneal cytology after curative resection for colorectal carcinoma." *Jpn J Clin Oncol* 33(1): 33-7.

- Yang, J. L., D. Seetoo, et al. (2000). "Urokinase-type plasminogen activator and its receptor in colorectal cancer: independent prognostic factors of metastasis and cancer-specific survival and potential therapeutic targets." *Int J Cancer* 89(5): 431-9.
- Zeng, Z. S., Y. Huang, et al. (1996). "Prediction of colorectal cancer relapse and survival via tissue RNA levels of matrix metalloproteinase-9." *J Clin Oncol* 14(12): 3133-40.
- Zhang, W., W. Tang, et al. (2008). "Positive KL-6 mucin expression combined with decreased membranous beta-catenin expression indicates worse prognosis in colorectal carcinoma." *Oncol Rep* 20(5): 1013-9.
- Zlobec, I., L. M. Terracciano, et al. (2008). "Local recurrence in mismatch repair-proficient colon cancer predicted by an infiltrative tumor border and lack of CD8+ tumor-infiltrating lymphocytes." *Clin Cancer Res* 14(12): 3792-7.

Appendix 3: Procure study - prospective registration

Patient data

National number^{REQ}:

Name^{REQ}: First name^{REQ}:

Date of birth (dd/mm/yyyy)^{REQ}:

Sex^{REQ}:

- ☐ Male
- ☐ Female

Zipcode of residence^{REQ}:

Registration number, provided by the data centre:

General practitioner (name, first name):

Hospital data

Contact person (can be a study nurse)

Contact details: Name:

Address:

Tel. Number:

Email address:

Name Hospital (1)^{REQ}:

Treatment (indicate treatments within the same hospital):

- ☐ Surgery: name surgeon(s)^{REQ}:
.....
- ☐ Preoperative staging.....
- ☐ Radiotherapy
- ☐ Chemotherapy
- ☐ Pathology report
- ☐ Follow-up: name responsible physician^{REQ}:
.....

Name Hospital (2):

Treatment:

- ☐ Preoperative staging.....
- ☐ Radiotherapy
- ☐ Chemotherapy
- ☐ Pathology report
- ☐ Follow-up: name responsible physician^{REQ}:
.....

Name Hospital (3):.....

Treatment:

- ☐ Preoperative staging.....
- ☐ Radiotherapy
- ☐ Chemotherapy
- ☐ Pathology report
- ☐ Follow-up: name responsible physician ^{REQ.} :.....

.....

OPERATIVE DATA ENTRY FORM

Registration number, provided by the data centre:.....

Name patient:.....First Name patient:.....

Date of Birth:...../...../.....

PART I: Pre-treatment data

1. Date of first consultation or hospitalisation for rectal cancer^{REQ}

(dd/mm/yyyy):.....

2. Synchronous cancer^{REQ} ?

☐ no

☐ yes

If yes:

a) organ(s):

☐ breast

☐ colon

☐ lung

☐ gynaecological tumour

☐ lymphoma

☐ other, please specify:

b) date of diagnosis: (dd/mm/yyyy):.....

c) cTNM stage: T..... N..... M.....

d) pTNM stage: T..... N..... M.....

3. Other cancer(s) in patient's past history^{REQ}?

☐ no

☐ yes :

If yes:

a) organ(s):

Tumour 1

☐ breast

☐ colon

☐ lung

☐ gynaecological tumour

☐ lymphoma

☐ other, please specify:

Tumour 2

☐ breast

☐ colon

☐ lung

☐ gynaecological tumour

☐ lymphoma

☐ other, please specify:

b) actual tumour activity ?tumour 1☐ yes☐ notumour 2☐ yes☐ no**4. Lower limit primary tumour:.....cm above the margo ani^{REQ}**

based on:

- ☐ rigid rectoscopy (to be preferred)
- ☐ coloscopy (during withdrawal of the coloscopy)

5. Characteristics of the primary tumour**Localisation^{REQ}:**

- ☐ Ventral
- ☐ Lateral left
- ☐ Lateral right
- ☐ Dorsal

Upper limitcm (if possible, in cm above the margo ani)**Clinical**

- ☐ Mobile
- ☐ Fixed
- ☐ Not palpable

6. Pretreatment staging procedures and clinical TNM (UICC 2002)

Check all staging procedures that were carried out.

Rx thorax ☐ yes ☐ no**US liver/abdomen** ☐ yes ☐ no**CT:**

- Thorax: ☐ yes ☐ no
- Abdomen/pelvis: ☐ yes ☐ no

If yes: cT:

cN:.....

cCRM lateral or circumferential margin:.....mm

MRI: ☐ yes ☐ noIf yes: cT:

cN:.....

cCRM lateral or circumferential margin:.....mm

involvement of the sphincters : ☐ yes ☐ no**TRUS:** ☐ yes ☐ noIf yes: cT:

cN:.....

involvement of the sphincters : ☐ yes ☐ no**PET** ☐ yes ☐ no**PET/CT** ☐ yes ☐ no**Other** ☐ yes ☐ noIf yes, please specify:

cM^{REQ}

- ☐ No metastasis
- ☐ Metastasis

Location:

- ☐ Non-mesorectal nodes (including external or common iliac nodes and retroperitoneal nodes above inf. mesenteric artery)
- ☐ Liver
- ☐ Peritoneum
- ☐ Lung
- ☐ Bone
- ☐ Other, please specify:

Based on:

- ☐ Rx thorax
- ☐ US liver/abdomen
- ☐ CT
- ☐ PET
- ☐ Other, please specify:

Summary cTNM stage^{REQ}: cT.....N.....M.....

7. CEA serum before treatment^{REQ}:

8. Colonoscopy*Total colonoscopy*^{REQ}

- ☐ Yes
 - simultaneous lesions?
 - ☐ No
 - ☐ Polyp
 - ☐ Carcinoma
 - ☐ Other
- ☐ No
 - Reason?
 - ☐ Tumour stenosis
 - ☐ Insufficient preparation
 - ☐ Intolerance of the patient
 - ☐ Technical reasons
 - ☐ Other

Biopsy of the tumour

- ☐ Yes
 - Date of biopsy^{REQ} (dd/mm/yyyy): ... / ... / ...
 - Result of the biopsy (specify):
 - ☐ Adenocarcinoma
 - ☐ Other:
- ☐ No

Complications

- ☐ No
- ☐ Yes

If yes:

- Oversedation ☐ yes ☐ no
- Bleeding ☐ yes ☐ no
- Perforation ☐ yes ☐ no
- Other ☐ yes ☐ no

9. Double contrast barium enema

- ☐ No
- ☐ Yes
- ☐ Barium ☐ Gastrografine
☐ Complete ☐ Incomplete (incompl. visualisation of the entire colon)

10. Virtual colonoscopy

- ☐ No
- ☐ Yes
- simultaneous lesions?
- ☐ No
☐ Polyp
☐ Carcinoma
☐ Other

11. Anorectal function before treatment**Continent?** ^{REQ}

- ☐ Yes
- ☐ No

Daily frequency of defaecation:.....

Use of drugs/medication for defaecation (incl. enema)

- ☐ No
- ☐ Yes

12. Urogenital function before treatment

Urinary function

Continent?

- ☐ Yes
- ☐ No

Sexual function

- ☐ Non active
- ☐ Active

Male:

- ☐ Normal
- ☐ dysfunction
- ☐ Not known

Female:

- ☐ Normal
- ☐ dysfunction
- ☐ Not known

13. Clinical restaging after neoadjuvant treatment (if applicable):

- Date of restaging (ddmmyyyy):...../...../.....
- Clinical response (choose 1 of the following)
 - ☐ No change in bulk
 - ☐ Increase in bulk
 - ☐ Reduction in bulk
 - ☐ Complete response
- Summary ycT (0,1,2,3,4) N (0,1,2) M (0,1,x)
- ycCRM : mm

OPERATIVE DATA ENTRY FORM

Registration number, provided by the data centre:.....

Name patient:.....First Name patient:.....

Date of Birth:...../...../.....

PART II: Operative data

1. WAS RADICAL RESECTION INDICATED BUT NOT PERFORMED? ^{REQ}

- ☐ No
- ☐ Yes

If Yes:

Reason(s):

- ☐ Patient unfit
- ☐ Patient refusal
- ☐ Advanced disease
- ☐ Other (specify):

2. TREATMENT OTHER THAN OR PRIOR TO RADICAL RESECTION ^{REQ}

- ☐ No
- ☐ Yes

If yes:

What treatment(s) was performed instead of or prior to radical resection?

- ☐ Abdominal exploration only
 - ☐ Laparotomy
 - ☐ laparoscopy
- ☐ Transanal laser or electrocautery
- ☐ Endoscopic stent
 - ☐ As definitive treatment: date (ddmmyyy):...../...../.....
 - ☐ As a bridge to surgery: date (ddmmyyy):...../...../.....
- ☐ Decompressive stoma

Date (ddmmyyy):...../...../.....

Approach

 - ☐ Laparotomy
 - ☐ without abdominal exploration
 - ☐ with abdominal exploration
 - ☐ no metastatic disease
 - ☐ metastatic disease
 - ☐ Laparoscopy
 - ☐ without abdominal exploration
 - ☐ with abdominal exploration
 - ☐ no metastatic disease
 - ☐ metastatic disease

Location

 - ☐ Ileum
 - ☐ Colon transversum
 - ☐ Sigmoid colon
 - ☐ other

Type

 - ☐ Loop
 - ☐ terminal

- ☐ “Local excision” (incl. endoscopic polypectomy and TEM)
 - Procedure
 - Endoscopic polypectomy: date (ddmmyyyy)...../...../.....
→ **Please fill in ‘local excision’ pathology report**
 - Local transanal excision: date (ddmmyyyy)...../...../.....
→ **Please fill in ‘local excision’ pathology report**
 - TEM (transanal endoscopic microsurgery): date (ddmmyyyy)...../...../.....
→ **Please fill in ‘local excision’ pathology report**
 - Intent
 - Curative treatment
 - Sampling (as an excisional biopsy)
- ☐ Neoadjuvant treatment
 - Short course radiotherapy with short interval to surgery
 - Short course radiotherapy with long interval to surgery
 - Long course chemoradiation with long interval to surgery
- ☐ Chemotherapy for cStage IV disease

3. RADICAL RESECTION

- ☐ No
- ☐ Yes

If Yes:

(Fill in all the following questions 3.1-3.13)

3.1. PLANNED type of radical resection^{REQ}:

- ☐ Hartmann
- ☐ APER
- ☐ Sphincter saving radical resection

3.2. Preoperative risk (factors of)

ASA (1-5)^{REQ}:

1. normal
2. mild systematic disease, normal activity
3. severe systematic disease, limited activity
4. life threatening disease, disabled
5. moribund

Hct^{REQ}:%

3.3. Preoperative Weight:.....kg

Height:.....cm

3.4. Date of surgery (dd/mm/yyyy)^{REQ}

3.5 Actual surgical training status:

- ☐ With trainer/instructor
- ☐ Self-training
- ☐ Peer to peer
- ☐ Trainer/instructor

3.6 Mode of surgery^{REQ}:

- ☐ Elective (operation at the time to suit both patient and surgeon)
- ☐ Scheduled (an early operation, but not immediately life-saving)
- ☐ Urgent (operation carried out within 24-hrs of admission)
- ☐ Emergency (immediate operation within 2 hours of admission or in conjunction with resuscitation)

3.7 Localisation of the primary tumour after anal investigation^{REQ}

- ☐ Ventral
- ☐ Lateral left
- ☐ Lateral right
- ☐ Dorsal
- ☐ no evidence of tumour

3.8 Lower limit of the primary tumour^{REQ}:..... cm above the margo ani based on:

- ☐ rigid rectoscopy (to be preferred)
- ☐ coloscopy (during withdrawal of the colonoscope)
- ☐ no evidence of tumour

3.9 Rectal irrigation at the start of the surgical procedure

- ☐ No
- ☐ Yes (specify the fluid).....

3.10 Surgical exploration**Approach:**

- ☐ Laparotomy
- ☐ Laparoscopy
- ☐ Converted laparoscopy: reason(s)
 - ☐ Adhesions
 - ☐ Bleeding
 - ☐ Bowel perforation
 - ☐ Other:

Ascites:

- ☐ No
- ☐ Yes

Cytology of ascites

- ☐ No
- ☐ Yes

Metastasis^{REQ}:

- ☐ No
- ☐ Exploration limited because of adherences
- ☐ Yes

<input type="checkbox"/> Liver	biopsy: <input type="checkbox"/> yes	<input type="checkbox"/> no
<input type="checkbox"/> Peritoneum	biopsy: <input type="checkbox"/> yes	<input type="checkbox"/> no
<input type="checkbox"/> Omentum	biopsy: <input type="checkbox"/> yes	<input type="checkbox"/> no
<input type="checkbox"/> Ovary	biopsy: <input type="checkbox"/> yes	<input type="checkbox"/> no
<input type="checkbox"/> Other	biopsy: <input type="checkbox"/> yes	<input type="checkbox"/> no

☐ Non-mesenterial lymph nodes

☐ Iliac

 biopsy: ☐ yes ☐ no

☐ Periaortic

 biopsy: ☐ yes ☐ no

☐ Hilus liver

 biopsy: ☐ yes ☐ no

☐ Celiac

 biopsy: ☐ yes ☐ no

Biopsy of metastasis
☐ No

☐ Yes (specify).....

Tumour:
Localisation of the tumour related to peritoneal reflection ^{REQ}
☐ Above

☐ At the level of

☐ Under

☐ Mobile

☐ Fixed

☐ Not palpable

Invasion into other organs (specify) ^{REQ}
☐ No

☐ Yes

☐ Pelvic wall

☐ Vagina

☐ Bladder

☐ Uterus

☐ Prostate

☐ Seminal vesicle(s)

☐ Ureter

☐ Colon

☐ Small bowel

Tumour complications before any mobilisation
☐ Peri-rectal abscess

☐ Stenosis or obstruction

☐ Free perforation

☐ Other:.....

3.11 Surgical resection
Approach ^{REQ}:

☐ Laparotomy

☐ Laparoscopy

☐ Converted laparoscopy (intention was to resect laparoscopically):

Reason(s) for conversion:

☐ Adhesions

☐ Bleeding

☐ Rectal perforation

☐ Other:

Procedure ^{REQ}:*Vascular ligatures* ^{REQ}:

- ☐ AMI
- ☐ VMI at the level of AMI
- ☐ VMI below the pancreas
- ☐ ARM
- ☐ Other, please specify:.....

Extent of the resection ^{REQ}:

‘en bloc’ resection of another organ?

- ☐ No
- ☐ Yes (specify):
 - ☐ Pelvic wall
 - ☐ Vagina
 - ☐ Bladder
 - ☐ Uterus (and ovaria)
 - ☐ Prostate
 - ☐ Seminal vesicle(s)
 - ☐ Ureter
 - ☐ Colon
 - ☐ Small bowel

deviation from the procedure of ‘en bloc’ resection? ^{REQ}

- ☐ No
- ☐ Yes (why?):.....

Non ‘en bloc’ resection of other organ ^{REQ}

- ☐ No
- ☐ Yes:
 - ☐ Ovaria
 - ☐ Liver
 - ☐ Peritoneum
 - ☐ Non-mesenterial node(s)
 - ☐ Other:.....

Perforation of the rectum? ^{REQ}

- ☐ No
- ☐ Yes

Complete resection of the sigmoid?

- ☐ Yes
- ☐ No

Distal level of resection (in case of reconstruction or Hartmann) ^{REQ}

- ☐ Rectum: cm above anal verge
- ☐ Anorectal (on top of the anal canal)
- ☐ Anal (intra-anal)

Technique used in case of sphincter saving resection ^{REQ}

- ☐ PME
- ☐ TME
- ☐ Conventional

Technique used in case of APER (abdominoperineal resection)

- ☐ perineal resection in supine position
- ☐ perineal resection in prone position

Autonomous nervous system

- ☐ Complete preservation
- ☐ Section hypogastric at the level of the promontorium
- ☐ Section left hypogastric
- ☐ Section right hypogastric
- ☐ Section pelvic plexus bilateral
- ☐ Section pelvic plexus left
- ☐ Section pelvic plexus right
- ☐ Not known
- ☐ Other

Peritoneal washing after resection, before or after reconstruction:

- ☐ No
- ☐ Yes (specify fluid):

*Which type of resection is clinically and surgically obtained^{REQ}:
(do not take the results of the pathology report into account)*

- ☐ R0
- ☐ R1 (frozen sections)
- ☐ R2
- ☐ Uncertain: why?
 - ☐ Locally
 - ☐ At distance

Problems during resection

- ☐ No
- ☐ Yes (specify):

3.12 Surgical reconstruction**Approach^{REQ}:**

- ☐ Laparotomy
- ☐ Laparoscopy (inc. lap-assisted)
- ☐ Converted laparoscopy

Complete mobilisation of the splenic flexure^{REQ}

- ☐ No
- ☐ Yes

Irrigation of the rectum stump before reanastomosis^{REQ}

- ☐ No
- ☐ Yes (specify fluid):

Type of reconstruction ^{REQ}

- ☐ APER (abdominoperineal excision; rectal amputation)
- ☐ Hartmann: distal transection level at cm above anal verge
- ☐ PME + High anterior resection (= colorectal anastomosis above peritoneal reflection)
- ☐ PME + Low anterior resection (= PME + colorectal anastomosis below peritoneal reflection)
- ☐ TME + Colon J pouch: length of pouch:cm
- ☐ TME + Colooplasty: length of incision for plasty:cm
- ☐ TME + side-to-end coloanal anastomosis
- ☐ TME + straight coloanal anastomosis
- ☐ TME) + Other (specify):.....
- ☐ Total excision of colon and rectum with ileal pouch-anal anastomosis
- ☐ Total excision of colon and rectum with definitive ileostomy
- ☐ Other (specify):

Distal anastomosis technique ^{REQ}:

- ☐ Stapled
- ☐ Manual

Derivative stoma after reconstruction (do not fill in in case of APER or Hartmann) ^{REQ}

- ☐ No
- ☐ Yes:
 - Place:
 - ☐ Colon
 - ☐ Ileum
 - ☐ Other:

Type:

- ☐ Loop
- ☐ Terminal

Reason(s)

- ☐ Routine (if done always with the type of reconstruction)
- ☐ Selective (specify reason(s))
 - ☐ ASA 3 or more
 - ☐ Difficult dissection
 - ☐ 1 L blood transfusion or more
 - ☐ Doubtful blood supply
 - ☐ Incomplete doughnut
 - ☐ Positive leak test
 - ☐ Poor bowel preparation
 - ☐ Radiotherapy
 - ☐ Other:.....

3.13 Intraoperative bloodtransfusion (not blood loss!) ^{REQ}:

- ☐ No
- ☐ Yes (specify volume of transfused packed cells): ml.
(1 unit PC = 400 ml)

OPERATIVE DATA ENTRY FORM

Registration number, provided by the data centre:.....

Name patient:.....First Name patient:.....

Date of Birth:...../...../.....

PART III: Post-operative data

1. Post-operative death ^{REQ}

- ☐ No
- ☐ Yes
- Date of death (dd/mm/yyyy):...../...../.....
- Cause of death:.....

2. Discharge date (dd/mm/yyyy) ^{REQ}:/...../.....

3. Discharge

- ☐ Home
- ☐ Other medical department (incl. geriatric)
- ☐ Revalidation centre
- ☐ Other:.....

4. Postoperative bloodtransfusion

- ☐ No
- ☐ Yes (specify volume of transfused packed cells):..... ml.
(1 unit PC = 400 ml)

5. Postoperative complications before discharge ^{REQ}

- ☐ No
- ☐ Yes

Medical

- ☐ Pneumonia
- ☐ Pulmonary embolism
- ☐ Myocardial infarction
- ☐ Cerebrovascular accident
- ☐ Catheter sepsis
- ☐ Renal insufficiency
- ☐ Urinary tract infection
- ☐ Pyelonephritis
- ☐ Deep venous thrombosis
- ☐ Other, please specify:.....

Surgical

(minor = no reintervention; major = reintervention under narcosis)

- ☐ Postoperative bleeding
 - ☐ Minor
 - ☐ Major

- ☐ Ileus (> 4D 'npo')
 - ☐ Minor
 - ☐ Major
- ☐ Urinary retention
- ☐ Abdominal wound infection
 - ☐ Minor
 - ☐ Major
- ☐ Perineal wound infection
 - ☐ Minor
 - ☐ Major
- ☐ Deep abscess
 - ☐ Minor
 - ☐ Major
- ☐ Leakage of the anastomosis
 - ☐ Minor
 - ☐ Major

Type of the reintervention(s)

(fill out numbers chronologically and add dates(dd/mm/yyyy) if applicable):

--	--	--	--	--

1. Derivative stoma construction date:.....
2. Dismantling of anastomosis (Hartmann) date:.....
3. Abdominal drainage date:.....
4. Transanal drainage date:.....
5. Other:.....date:.....

- ☐ Complication of the stoma
 - ☐ Minor
 - ☐ Major

Type of complication (with influence on hospitalisation):

.....

.....

Type of re-intervention (specify)

.....

.....

RADIOTHERAPY DATA ENTRY FORM

Registration number, provided by the data centre:.....

Name patient:.....First Name patient:.....

Date of Birth:...../...../.....

Treatment^{REQ}

- ☐ Preoperative radiotherapy
- ☐ Postoperative radiotherapy

Concomitant chemotherapy^{REQ}

- ☐ No
- ☐ Yes

If Yes:

5-FU based ?

- ☐ yes
- ☐ no

Treatment position^{REQ}:

- ☐ Supine
- ☐ Prone

Belly board^{REQ}:

- ☐ Yes
- ☐ No

Planned irradiation regimen: x Gy

Date of first irradiation (dd/mm/yyyy)^{REQ}:/...../.....

Date of last irradiation (dd/mm/yyyy)^{REQ}:/...../.....

Number of fractions^{REQ}

Radiation compliance: treatment interruption of more than five working days^{REQ}:

- ☐ No
- ☐ Yes

Reason for treatment interruption of more than five working days:

- ☐ Toxicity
- ☐ Machine break down
- ☐ Other:

Total dose given at ICRU reference point^{REQ}:

Custom shielding^{REQ}:

- ☐ MLC
- ☐ Blocks
- ☐ No

The photon energy used was^{REQ}:

- ☐ Co⁶⁰
- ☐MV

Number of beams used:

Technique used ^{REQ}:

- ☐ 2D
- ☐ 3D CRT
- ☐ IMRT
- ☐ IMAT (including VMAT/RapidARC)
- ☐ HT (helical tomotherapy)

Only for 2D planning (simulation)

Field sizes if 2D: F1:

F2:

Applicable for CT-based planning:

Total volume irradiated to 95% ^{REQ}:

PTV: Mean dose ^{REQ}:

Median dose:

Maximum dose:

Minimum dose:

PTV BOOST:

☐ No ^{REQ}

☐ Yes ^{REQ}

If yes:

Mean dose ^{REQ}:

Median dose:

Maximum dose:

Minimum dose:

Organs at risk (OARs)

- Small bowel absolute volume (cc) > 15 Gy: cc

- Bladder volume (%) > 40 Gy: %

- Femoral heads combined volume (%) > 40 Gy: %

[illegible]

PATHOLOGY REPORT CHECKLIST AFTER LOCAL EXCISION (incl. polypectomy, transanal resection, TEMS) ^{REQ}

Patient's name: Patient's first name: Date of birth:	Registration number: Hospital/Laboratory: Pre-operative treatment (no/yes+what):
RECTAL CANCER: Distance from anal verge ... cm cTNM staging:	
TYPE OF INTERVENTION <input type="checkbox"/> Endoscopic polypectomy <input type="checkbox"/> Transanal local excision <input type="checkbox"/> TEMS	TUMOR LOCATION <input type="checkbox"/> ventral <input type="checkbox"/> above peritoneal reflection <input type="checkbox"/> lateral <input type="checkbox"/> below peritoneal reflection <input type="checkbox"/> dorsal <input type="checkbox"/> Multifocal: if second location, please use separate sheet
MACROSCOPIC EXAMINATION <input type="checkbox"/> fresh <input type="checkbox"/> fixed Photos of the fresh specimen : yes – no Number of fragments Dimensions of resected specimen:x..... xcm Distance tumor – resection margin: proximal:cm distal:cm lateral left:cm lateral right:cm depth:cm Rectal tumor <input type="checkbox"/> exophytic <input type="checkbox"/> ulcerating <input type="checkbox"/> infiltrating <input type="checkbox"/> flat Additional samples: <input type="checkbox"/> frozen <input type="checkbox"/> other fixation	HISTOLOGIC EXAMINATION <input type="checkbox"/> Adenocarcinoma <input type="checkbox"/> well <input type="checkbox"/> low grade <input type="checkbox"/> moderate <input type="checkbox"/> high grade <input type="checkbox"/> poorly differentiated <input type="checkbox"/> undifferentiated <input type="checkbox"/> Other: Depth of invasion <input type="checkbox"/> T0 <input type="checkbox"/> Tis: intra-mucosal or intra-epithelial (not beyond muscularis mucosae) - m1 - m2 - m3 <input type="checkbox"/> T1: limited to submucosa - sm1 - sm2 - sm3 <input type="checkbox"/> T2: limited to muscularis propria <input type="checkbox"/> T3 <input type="checkbox"/> T4 Surgical resection : Margins: Proximal: <input type="checkbox"/> free.....mm <input type="checkbox"/> invaded Distal: <input type="checkbox"/> free.....mm <input type="checkbox"/> invaded Lateral left: <input type="checkbox"/> free.....mm <input type="checkbox"/> invaded Lateral right: <input type="checkbox"/> free.....mm <input type="checkbox"/> invaded Depth: <input type="checkbox"/> free.....mm <input type="checkbox"/> invaded Extension: <input type="checkbox"/> lymphovascular invasion: o yes o no <input type="checkbox"/> number of lymph nodes found: <input type="checkbox"/> number of invaded lymph nodes:
RECTAL CANCER <input type="checkbox"/> pTNM <input type="checkbox"/> YpTNM Other classification:	<input type="checkbox"/> T0 <input type="checkbox"/> Tis <input type="checkbox"/> T1 <input type="checkbox"/> T2 <input type="checkbox"/> T3 <input type="checkbox"/> -m1 <input type="checkbox"/> -sm1 <input type="checkbox"/> -m2 <input type="checkbox"/> -sm2 <input type="checkbox"/> -m3 <input type="checkbox"/> -sm3 <hr/> <input type="checkbox"/> Nx <input type="checkbox"/> N+ N0
Signature : Date :	

CHEMOTHERAPY DATA ENTRY FORM

Registration number, provided by the data centre:.....,

Name patient:.....First Name patient:.....

Date of Birth:...../...../.....

To be filled out at the start of chemotherapy

Treatment^{REQ}

- ☐ Neoadjuvant chemotherapy
 - ☐ with radiotherapy
 - ☐ without radiotherapy

- ☐ Adjuvant chemotherapy
 - ☐ with radiotherapy
 - ☐ without radiotherapy

- ☐ Palliative chemotherapy
 - ☐ NO surgery planned
 - ☐ because of the extent of the disease
 - ☐ because of age and/or comorbidities
 - ☐ because of patient refusal
 - ☐ other:
 - ☐ BEFORE planned surgery for primary, metastatic disease or both (in any sequence)
 - ☐ surgery POTENTIALLY planned during/after palliative chemotherapy
 - ☐ AFTER resectional surgery of metastasis with following status:
 - ☐ R 0 ("no residual disease")
 - ☐ R 1 (at least one resection with a positive margin)
 - ☐ R 2 (at least one metastasis present)

CHEMOTHERAPY DATA ENTRY FORM

Registration number, provided by the data centre:.....

Name patient:.....First Name patient:.....

Date of Birth:...../...../.....

To be filled out at the end of chemotherapy

Weight:kg^{REQ} **Length:**m^{REQ}

Type of medication (dose expressed per m²)^{REQ}

1. Neoadjuvant chemotherapy with radiotherapy

- ☐ 5 FU: - schedule
 - ☐ bolus
 - ☐ continuous infusion
 - planned dose 5FU: mg/m²
 - global administered dose 5FU: mg
 - period (date) from/...../..... till...../...../.....,
- ☐ oral fluoropyrimidines ☐ capecitabine
 - ☐ other.....
 - planned dose: mg/m²
 - global administered dose: mg
 - period (date) from/...../..... till/...../.....,
- ☐ other:.....
 - schedule:
 - planned dose: mg/m²
 - global administered dose: mg
 - period (date) from...../...../..... till/...../.....

Dose reduction performed^{REQ}

- ☐ Yes
- ☐ No

Toxicity

- ☐ hospitalisation needed for toxicity exclusively due to chemotherapy
 - ☐ Yes
 - ☐ No (other treatment modality contributed also)
- ☐ leading to stopping chemotherapy
- ☐ leading to temporarily interrupting chemotherapy
- ☐ leading to dose reduction

Type of adverse events during chemotherapy or chemoradiotherapy (mention only grade 3-4 (severe) adverse events to be evaluated according to the NCI-CTC version 3.0 criteria)^{REQ}

- diarrhea: ☐ yes ☐ no if yes: grade 3 / 4
- nausea: ☐ yes ☐ no if yes: grade 3 / 4
- vomiting: ☐ yes ☐ no if yes: grade 3 / 4
- anorexia: ☐ yes ☐ no if yes: grade 3 / 4
- neutropenia: ☐ yes ☐ no if yes: grade 3 / 4
- neutropenic fever or infection: ☐ yes ☐ no
- anemia: ☐ yes ☐ no if yes: grade 3 / 4
- thrombocytopenia: ☐ yes ☐ no if yes: grade 3 / 4
- stomatitis: ☐ yes ☐ no if yes: grade 3 / 4
- neurotoxicity: ☐ yes ☐ no if yes: grade 3 / 4
- hand-foot syndrome: ☐ yes ☐ no if yes: grade 3 / 4
- other (specify):.....
-

2. Preoperative chemotherapy without radiotherapy

- ☐ 5 FU: - schedule
 - ☐ bolus
 - ☐ continuous infusion
 - planned dose 5FU: mg/m²
 - global administered dose 5FU: mg
 - period (date) from/...../..... till/...../.....
- ☐ oral fluoropyrimidines ☐ capecitabine
 - ☐ other.....
 - planned dose: : mg/m²
 - global administered dose: mg
 - period (date) from/...../..... till...../...../.....
- ☐ other:.....
 - schedule:
 - planned dose : mg/m²
 - global administered dose: mg
 - period (date) from/...../..... till/...../.....

Dose reduction performed

- ☐ Yes
- ☐ No

Toxicity^{REQ}

- ☐ hospitalisation needed for toxicity exclusively due to chemotherapy
 - ☐ Yes
 - ☐ No (other treatment modality contributed also)
- ☐ leading to stopping chemotherapy
- ☐ leading to temporarily interrupting chemotherapy
- ☐ leading to dose reduction

Type of adverse events during chemotherapy or chemoradiotherapy (mention only grade 3-4 (severe) adverse events to be evaluated according to the NCI-CTC version 3.0 criteria)^{REQ}

- diarrhea: ☐ yes ☐ no if yes: grade 3 / 4
- nausea: ☐ yes ☐ no if yes: grade 3 / 4
- vomiting: ☐ yes ☐ no if yes: grade 3 / 4
- anorexia: ☐ yes ☐ no if yes: grade 3 / 4
- neutropenia: ☐ yes ☐ no if yes: grade 3 / 4
- neutropenic fever or infection: ☐ yes ☐ no
- anemia: ☐ yes ☐ no if yes: grade 3 / 4
- thrombocytopenia: ☐ yes ☐ no if yes: grade 3 / 4
- stomatitis: ☐ yes ☐ no if yes: grade 3 / 4
- neurotoxicity: ☐ yes ☐ no if yes: grade 3 / 4
- hand-foot syndrome: ☐ yes ☐ no if yes: grade 3 / 4
- other (specify):.....
.....

3. Adjuvant chemotherapy with radiotherapy

- ☐ 5 FU: - schedule
 - ☐ bolus
 - ☐ continuous infusion
 - planned dose 5FU:mg/m2
 - global administered dose 5FU: mg
 - period (date) from/...../..... till...../...../.....
- ☐ oral fluoropyrimidines ☐ capecitabine
 - ☐ other.....
 - planned dose:mg/m2
 - global administered dose: mg
 - period (date) from/...../..... till/...../.....
- ☐ other:.....
 - schedule:
 - planned dose:.....mg/m2
 - global administered dose: mg
 - period (date) from/...../..... till/...../.....

Dose reduction performed

- ☐ Yes
- ☐ No

Toxicity^{REQ}

- ☐ hospitalisation needed for toxicity exclusively due to chemotherapy
 - ☐ Yes
 - ☐ No (other treatment modality contributed also)
- ☐ leading to stopping chemotherapy
- ☐ leading to temporarily interrupting chemotherapy
- ☐ leading to dose reduction

Type of adverse events during chemotherapy or chemoradiotherapy (mention only grade 3-4 (severe) adverse events to be evaluated according to the NCI-CTC version 3.0 criteria)^{REQ}

- diarrhea: ☐ yes ☐ no if yes: grade 3 / 4
- nausea: ☐ yes ☐ no if yes: grade 3 / 4
- vomiting: ☐ yes ☐ no if yes: grade 3 / 4
- anorexia: ☐ yes ☐ no if yes: grade 3 / 4
- neutropenia: ☐ yes ☐ no if yes: grade 3 / 4
- neutropenic fever or infection: ☐ yes ☐ no
- anemia: ☐ yes ☐ no if yes: grade 3 / 4
- thrombocytopenia: ☐ yes ☐ no if yes: grade 3 / 4
- stomatitis: ☐ yes ☐ no if yes: grade 3 / 4
- neurotoxicity: ☐ yes ☐ no if yes: grade 3 / 4
- hand-foot syndrome: ☐ yes ☐ no if yes: grade 3 / 4
- other (specify):.....
.....

4. Adjuvant chemotherapy without radiotherapy

- ☐ 5 FU: - schedule
 - ☐ bolus
 - ☐ continuous infusion
 - planned dose 5FU: mg/m²
 - global administered dose 5FU: mg
 - period (date) from/...../..... till/...../.....
- ☐ oral fluoropyrimidines ☐ capecitabine
 - ☐ other.....
 - planned dose: mg/m²
 - global administered dose: mg
 - period (date) from/...../..... till/...../.....
- ☐ FOLFOX (5FU + Oxaliplatin)
 - 5FU:
 - schedule
 - ☐ bolus
 - ☐ continuous infusion
 - planned dose 5FU: mg/m²
 - global administered dose 5FU: mg
 - period (date) from/...../..... till/...../.....
 - Oxaliplatin
 - planned dose oxa: mg/m²
 - global administered dose oxa: mg
 - period (date) from/...../..... till/...../.....
- ☐ XELOX
 - Capecitabine cfr supra
 - planned dose: mg/m²
 - global administered dose: mg
 - period (date) from/...../..... till/...../.....

5. Palliative chemotherapy (please, use a new form with patient's name or national number for 2nd line etc.)

Regimen	1 st line	2 nd line	3 rd line	4 th line
Oral fluoropyrimidine				
LV5FU2 (De Gramont)				
Folfox				
Folfiri				
Xelox				
Oral fluoropyrimidine + bevacizumab				
LV5FU2 (De Gramont) + bevacizumab				
Folfox + bevacizumab				
Folfiri + bevacizumab				
Xelox + bevacizumab				
Cetuximab + irinotecan				
Mitomycine + 5FU or capecitabine				
Other:				

Dose reduction performed

- ☐ Yes: percentage:%
- ☐ No

Toxicity^{REQ}

- ☐ hospitalisation needed for toxicity exclusively due to chemotherapy
 - ☐ Yes
 - ☐ No (other treatment modality contributed also)
- ☐ leading to stopping chemotherapy
- ☐ leading to temporarily interrupting chemotherapy
- ☐ leading to dose reduction

Type of adverse events during chemotherapy or chemoradiotherapy (mention only grade 3-4 (severe) adverse events to be evaluated according to the NCI-CTC version 3.0 criteria)^{REQ}

- diarrhea: ☐ yes ☐ no if yes: grade 3 / 4
- nausea: ☐ yes ☐ no if yes: grade 3 / 4
- vomiting: ☐ yes ☐ no if yes: grade 3 / 4
- anorexia: ☐ yes ☐ no if yes: grade 3 / 4
- neutropenia: ☐ yes ☐ no if yes: grade 3 / 4
- neutropenic fever or infection: ☐ yes ☐ no
- anemia: ☐ yes ☐ no if yes: grade 3 / 4
- thrombocytopenia: ☐ yes ☐ no if yes: grade 3 / 4
- stomatitis: ☐ yes ☐ no if yes: grade 3 / 4
- neurotoxicity: ☐ yes ☐ no if yes: grade 3 / 4
- hypertension: ☐ yes ☐ no if yes: grade 3 / 4
- proteinuria: ☐ yes ☐ no if yes: grade 3 / 4
- other (specify):.....
.....

Is the patient dead?

- ☐ No
- ☐ Yes

If yes:

- death due to chemotherapy alone
 - ☐ Yes
 - ☐ No
- death due to chemoradiotherapy
 - ☐ Yes
 - ☐ No

FOLLOW-UP DATA ENTRY FORM

Registration number, provided by the data centre:.....

Name patient:.....First Name patient:.....

Date of Birth:...../...../.....

Fill in one form for each follow-up period (i.e. every 6 months regarding to the initial incidence date (with incidence date 1. First histological/cytological confirmation 2. Clinical evaluation/hospitalization 3. First Treatment). Indicate the period that is applicable (please choose the period that is closest to the real time-interval) and fill in till 5 year or until an event occurs, i.e. until recurrent local disease or metachronous distant disease or death.

Please, continue follow-up until death for patients with primary cStage IV or pStage IV.

Follow-up time interval (period) ^{REQ}

- ☐ 6 mo
- ☐ 12 mo
- ☐ 18 mo
- ☐ 24 mo
- ☐ 30 mo
- ☐ 36 mo
- ☐ 42 mo
- ☐ 48 mo
- ☐ 54 mo
- ☐ 60 mo
- ☐

1. Did the patient receive chemotherapy in this 6 mo. interval (period) ^{REQ}

- ☐ No
- ☐ Yes

2. WHO Performance score

- ☐ 0 = normal activity
- ☐ 1 = symptomatic but ambulatory
- ☐ 2 = bedridden <50% per day
- ☐ 3 = bedridden >50% per day
- ☐ 5 = 100% bedridden

3. LATE COMPLICATIONS OF RADIO CHEMOTHERAPY ^{REQ}:

- ☐ No
- ☐ Yes

If yes: (RTOG/EORTC grading 0-5; fill in max. grade per item)

- ☐ Skin:
- ☐ GI (small/large bowel):
- ☐ Bladder:
- ☐ Ureter:
- ☐ Nerves:
- ☐ Other (specify):

4. STOMA ^{REQ}:

- ☐ Not applicable (never had)
- ☐ Present
- ☐ Closed
- Date closure of stoma (dd/mm/yyyy):/...../.....
- (if applicable in this follow-up period)

5. ANORECTAL FUNCTION

- ☐ Continent ^{REQ}
 - ☐ Yes
 - ☐ No
 - ☐ Not applicable (APER, Hartmann, Derivative stoma)
- ☐ Defecation
 - Frequency per day or per week:/day or/week
- ☐ Medication related to defecation (incl. enemas)
 - ☐ No
 - ☐ Yes (specify):

6. UROGENITAL FUNCTION as compared with 6 mo. ago

Urinary function

- ☐ Idem
- ☐ Better
- ☐ Worse

Specific treatment:

- ☐ No
- ☐ Yes (specify):

Sexual function

- ☐ Not active
- ☐ Active
 - If active:
 - ☐ Idem
 - ☐ Better
 - ☐ Worse

Specific treatment:

- ☐ No
- ☐ Yes (specify):

7. LATE MEDICAL OR SURGICAL COMPLICATIONS ^{REQ}*during the preceding 6 mo*

- ☐ Type (specify):
- ☐ Date of diagnosis (dd/mm/yyyy):
- ☐ Treatment (specify briefly):
- ☐ Comment:

8. EXAMINATIONS DONE AT THE OCCASION OF THIS FOLLOW UP ^{REQ}**Indicate what was done**

- ☐ Colonoscopy : if yes, date (dd/mm/yyyy):...../...../.....
- ☐ RX thorax

- ☐ US liver
- ☐ CT abdomen/pelvis
- ☐ CT thorax
- ☐ CT thorax/abdomen
- ☐ PET
- ☐ PET/CT
- ☐ CEA
- ☐ Other(s):

9. NEW PRIMARY TUMOUR ^{REQ}

- ☐ No
 - ☐ Yes
- Date of diagnosis (dd/mm/yyyy):/...../.....

Localisation:

- ☐ Colon
- ☐ Other (specify):

Treatment:

- ☐ None
- ☐ Chemotherapy
- ☐ Radiotherapy
- ☐ Radiochemotherapy
- ☐ Surgery
- ☐ Other:
- ☐ Comment:

10. LOCAL RECURRENCE ^{REQ}:

- ☐ No
- ☐ Yes

If yes, this is the final update for the PROCARE registry, but fill in the following

- ☐ Date of diagnosis (dd/mm/yyyy):/...../.....
- ☐ Localisation(s) multiple selection possible:
 - ☐ Laparotomy wound
 - ☐ Trocar (port) site(s)
 - ☐ Perineal wound
 - ☐ Small pelvis (excl. external or common iliac lymph nodes)
 - ☐ External or common iliac nodes
 - ☐ Other:
- ☐ Diagnostic proof (check):
 - ☐ clinical
 - ☐ endoscopy
 - ☐ TRUS
 - ☐ CT
 - ☐ MRI
 - ☐ Biopsy/cyto
 - ☐ CEA
 - ☐ Other:

- ☐ Treatment:
- ☐ None
 - ☐ Chemotherapy
 - ☐ Radiotherapy
 - ☐ Radiochemotherapy
 - ☐ Surgery
 - ☐ Palliative measures
 - ☐ Other.....
 - ☐ Comment.....

11. METACHRONOUS DISTANT METASTASIS

(metachronous = diagnosed more than 6 months after incidence date i.e. date of diagnosis of rectal cancer) ^{REQ}

- ☐ no
- ☐ yes

If yes, this is the final update for the PROCARE registry but fill in the following

- ☐ Date of diagnosis (dd/mm/yyyy):/...../.....
- ☐ Localisation(s): multiple selection possible
 - ☐ Liver
 - ☐ Lung
 - ☐ Peritoneum
 - ☐ Para-aortic nodes
 - ☐ Bone
 - ☐ Other, please specify:
- ☐ Diagnostic proof (check):
 - ☐ Clinical
 - ☐ CEA
 - ☐ US
 - ☐ RX thorax
 - ☐ CT
 - ☐ MRI
 - ☐ Bone scan
 - ☐ PET
 - ☐ Biopsy/cyto
 - ☐ Other:
- ☐ Treatment:
 - ☐ None
 - ☐ Chemotherapy
 - ☐ Radiotherapy
 - ☐ Radiochemotherapy
 - ☐ Surgery
 - ☐ Palliative measures
 - ☐ Other.....
 - ☐ Comment.....

12. DEATH

- ☐ Date (dd/mm/yyyy):/...../.....

- ☐ Cause (check) **REQ.**
 - ☐ Cancer related
 - ☐ Death related to registered primary
 - ☐ Death related to another primary
 - ☐ Death related to metastases from unknown origin
 - ☐ Unknown
 - ☐ Other (specify):

Appendix 4: QCIs discussed in the PROCARE consensus (July 2010)

QCIS DISCUSSED IN THE PROCARE CONSENSUS (JULY 2010)	196
1 QUALITY OF CARE INDICATORS PER DOMAIN	197
1.1 GENERAL QUALITY INDICATORS	197
1.1.1 Description of the QCIs	197
1.2 DIAGNOSIS AND STAGING	199
1.2.1 Description of the QCIs	199
1.3 NEOADJUVANT TREATMENT	203
1.3.1 Description of the QCIs	203
1.4 SURGERY	206
1.4.1 Description of the QCIs	206
1.5 ADJUVANT TREATMENT	211
1.5.1 Description of the QCIs	211
1.6 PALLIATIVE TREATMENT	213
1.6.1 Description of the QCI	213
1.7 FOLLOW-UP	214
1.7.1 Description of the QCIs	214
1.8 HISTOPATHOLOGIC EXAMINATION	215
1.8.1 Description of the QCIs	215
2 OUTCOME-SPECIFIC QUALITY OF CARE INDICATORS	217
3 PROCESS-SPECIFIC QUALITY OF CARE INDICATORS	218

1 QUALITY OF CARE INDICATORS PER DOMAIN

1.1 FOREWORD

In this section we list QCI definitions as provided by the PROCARE consensus meeting. For operational reasons more technical definitions were sometimes needed. These can be found in Appendix 6 under the heading “Working definition”.

1.2 GENERAL QUALITY INDICATORS

Table 1: List of QCIs in the domain of general quality indicators

Code	Description	Type
1111	Overall 5-year survival by stage	Outcome
1112	Disease-specific 5-year survival by stage	Outcome
new	Relative survival	Outcome
1113	Proportion of patients with local recurrence	Outcome
new	Disease-free survival	Outcome

1.2.1 Description of the QCIs

1.2.1.1 *Overall 5-year survival by stage (KCE 2008 QCI 1111; outcome indicator)*

N: Number of patients in denominator that survived 1-5 years

D: Number of patients for whom the national registry number is known and have a follow-up of 1 -5 years, respectively. Survival status was obtained through cross-link with the Crossroads Bank for Social Security (CBSS).

This QCI is called observed survival in PROCARE feedback. Survival curves were calculated using the Kaplan Meier method.

1.2.1.2 *Disease-specific 5-year survival by stage (KCE 2008 QCI 1112; outcome indicator)*

The percentage of people in a study or treatment group who have not died from rectal cancer in a defined period of time. The time period begins at the incidence date. Date of incidence is defined by the date of pathological diagnosis (biopsy), if missing by the date of first consultation or hospitalization, if still missing by the date of first treatment (any type).

Patients who died without rectal cancer (LR or metastasis) are censored.

1.2.1.3 *Relative survival (new QCI; outcome indicator)*

The relative survival is the ratio of observed survival in a population to the expected survival rate. It estimates the chance that a patient will survive a set number of years after a cancer diagnosis. It is calculated to exclude the chance of death from diseases other than the cancer and shows whether or not that specific disease shortens a person's life.

If reliable information on cause of death is available, it is preferable to use the 'adjusted rate', i.e. disease (rectal cancer)-specific survival. This is particularly true when the series is small or when the patients are largely drawn from a particular segment of the population (e.g. socioeconomic segment).

1.2.1.4 *Proportion of patients with local recurrence (KCE 2008 QCI 1113; outcome indicator)*

N: Number of patients in denominator who developed a local recurrence at 1-5 year

D: Number of (y)pStage 0-III patients with R0 resection who have a follow-up of 1-5 years, respectively.

Local recurrence rate curves are calculated using the Kaplan Meier method.

1.2.1.5 *Disease-free survival (new QCI; outcome indicator)*

N: Number of patients in denominator who did not develop a local recurrence and/or distant metastasis at 1-5 year of follow-up.

D: Number of (y)pStage 0-III patients with R0 resection who have a follow-up of 1-5 years, respectively.

Disease-free survival rate curves were calculated using the Kaplan Meier method.

1.3 DIAGNOSIS AND STAGING

Table 2: List of QCI in the domain of diagnosis and staging

Code	Description	Type
1211	Proportion of patients with a documented distance from the anal verge	Process
1212	Proportion of patients in whom a CT of the abdomen and RX or CT thorax was performed before any treatment	Process
1213	Proportion of patients in whom a CEA was performed before any treatment	Process
1214	Proportion of patients undergoing elective surgery that had preoperative complete large bowel-imaging	Process
1215	Proportion of patients in whom a TRUS and pelvic CT and/or pelvic MRI was performed before any treatment	Process
1216	Proportion of patients with cStage II-III RC that have a reported cCRM	Process
1217	Time between first histopathologic diagnosis and first treatment	Process
new	Accuracy of cM0 staging	Process
new	Accuracy of cT/cN staging if no or short radiotherapy (separately presented in 2 tables)	Process
new	Use of TRUS in cT1/cT2	Process
new	Use of MRI in cStage II or III	Process

1.3.1 Description of the QCIs

1.3.1.1 *Proportion of patients with a documented distance from the anal verge (KCE 2008 QCI 1211; process indicator)*

N: Number of patients in denominator for whom lower limit of the tumour is known (see definition lower limit of tumour)

D: Number of registered patients

Priority sequence to determine lower limit:

1. pretreatment rectoscopy,
2. pretreatment colonoscopy,
3. rectoscopy or colonoscopy at surgery.

Table 3: Level of tumour (lower limit determined by distance from anal verge)

Lower limit tumour (LL)	Level tumour
≤ 5 cm	Low
>5 - ≤ 10 cm	Mid
>10 cm	High

For patients with long course neoadjuvant radiotherapy the pretreatment lower limit is taken as lower limit of the tumour. If no lower limit is available before neoadjuvant treatment, the lower limit measured at surgery is taken as lower limit of the tumour.

For patients who received neoadjuvant treatment but for whom it is not known whether they received short or long course radiotherapy, the lowest limit of either the pretreatment or the lower limit at surgery is taken.

1.3.1.2 *Proportion of patients in whom a CT of the abdomen and RX or CT thorax was performed before any treatment (KCE 2008 QCI 1212; process indicator)*

N: Number of patients in denominator in whom an abdominal CT and (rx thorax or CT thorax) was performed before any treatment

D: Number of registered patients with elective or scheduled surgery after August 1st 2008.

Until now not used for PROCARE feedback because the use of CT may be underestimated in patients registered using forms dating prior to August 1st 2008 (related to the structure and formulation of the early forms).

1.3.1.3 *Proportion of patients in whom a CEA was performed before any treatment (KCE 2008 QCI 1213; process indicator)*

N: Number of patients in denominator for whom CEA serum level before treatment is reported

D: Number of registered patients

1.3.1.4 *Proportion of patients undergoing elective surgery that had preoperative complete large bowel-imaging (KCE 2008 QCI 1214; process indicator)*

N: Number of patients in denominator who underwent a total colonoscopy or a complete double contrast enema or virtual colonoscopy

D: Number of patients treated with elective or scheduled surgery.

1.3.1.5 *Proportion of patients in whom a TRUS and pelvic CT and/or pelvic MRI was performed before any treatment (KCE QCI 1215; process indicator)*

N: Number of patients in whom cT or cN were based on TRUS and at least one of the two following:

- pelvic CT

- pelvic MRI

D: Number of registered patients with rectal cancer of any stage

CAUTION: may be underestimated in patients registered using forms dating prior to August 1st 2008.

1.3.1.6 *Proportion of patients with cStage II-III RC that have a reported cCRM (KCE QCI 1216; process indicator)*

N: Number of patients in denominator for whom cCRM is reported

D: Number of patients with cStage II-III treated with radical surgical resection.

1.3.1.7 *Time between first histopathologic diagnosis and first treatment (KCE QCI 1217; process indicator)*

For the patients treated by surgery and/or radiotherapy and/or chemotherapy, the time interval in days is computed between the date of pathologic diagnosis, if available, otherwise the date of first contact/hospitalization, and the date of first treatment.

1.3.1.8 *Accuracy of cM0 staging (new QCI; process indicator)*

N: Patients in denominator in whom no metastatic disease was diagnosed within 3months following the date of first treatment (any type).

D: All patients with cStage I-III and for whom a 1 year follow-up is available.

1.3.1.9 *Accuracy of cT/cN staging if no or short radiotherapy (separately presented in 2 tables) (new QCI; process indicator)*

For patients who did not receive neoadjuvant long course radio(chemo)therapy, the (y)pT/(y)pN is shown related to the cT/cN for these patients.

D: All patients with TRUS/CT/MRI with no or short neoadjuvant radiotherapy (without long R(C)T) and for whom the pT and pN is known and for whom the cT and cN is known (excluding patients with c and/or pTx and/or c and/or pNx

1.3.1.10 *Use of TRUS in cT1/cT2 (new QCI; process indicator)*

N: Number of patients in denominator in whom cT was based on TRUS

D: Number of patients with cT1 or cT2 rectal cancer registered after August 1st 2008

CAUTION: the use of TRUS may be underestimated in patients registered using forms dating prior to August 1st 2008.

1.3.1.11 Use of MRI in cStage II or III (new QCI; process indicator)

N: Number of patients in denominator in whom cT was based on MRI

D: Number of patients with cStage II or III rectal cancer based on any imaging technique registered after August 1st 2008.

CAUTION: the use of MRI may be underestimated in patients registered using forms dating prior to August 1st 2008.

1.4 NEOADJUVANT TREATMENT

Definition:

- Short course regimen are 5 x 5, 10 or 13 x 3 Gy (always without chemotherapy).
- Long course regimen are 25 or more x 1.8 Gy (with or without chemotherapy).

Table 4: List of QCI's in the domain of neoadjuvant treatment

Code	Description	Type
new	Proportion of cStage II-III patients that received a neoadjuvant pelvic RT	Process
new	Proportion of patients with cCRM \leq 2 mm on MRI/CT that received long course neoadjuvant radio(chemo)therapy	Process
new	Proportion of patients with cStage I that received neoadjuvant radio(chemo)therapy	Process
1224	Proportion of cStage II-III patients treated with neoadjuvant 5-FU based chemoradiation, that received a continuous infusion of 5-FU	Process
1225	Proportion of cStage II-III patients treated with a long course of preoperative pelvic RT or chemoradiation, that completed this neoadjuvant treatment within the planned timing	Process
1226	Proportion of cStage II-III patients treated with a long course of preoperative pelvic RT or chemoradiation, that was operated 4 to 12 weeks after completion of the (chemo)radiation	Process
1227	Rate of acute grade 4 radio(chemo)therapy-related complications	Process

1.4.1 Description of the QCI's

1.4.1.1 *Proportion of cStage II-III patients that received a neoadjuvant pelvic RT (new QCI; process indicator)*

For high rectal cancer (> 10 cm)

N: Number of patients in denominator who received neoadjuvant **R(C)T**

D: Number of patients in cStage II or III, treated with radical surgical resection with tumour in upper third

For mid rectal cancer (>5 - 10 cm)

N: Number of patients in denominator who received neoadjuvant **R(C)T**

D: Number of patients in cStage II or III, treated with radical surgical resection with tumour in middle third

For low rectal cancer (≤ 5 cm)

N: Number of patients in denominator who received neoadjuvant treatment

D: Number of patients in cStage II or III, treated with radical surgical resection with tumour in lower third

1.4.1.2 Proportion of patients with cCRM ≤ 2 mm on MRI/CT that received long course neoadjuvant radio(chemo)therapy (new QCI; process indicator)

N: Number of patients in denominator who received long course neoadjuvant radio(chemo)therapy

D: Number of patients treated with radical surgical resection and for whom cCRM is ≤ 2 mm

1.4.1.3 Proportion of patients with cStage I that received neoadjuvant radio(chemo)therapy (new QCI; process indicator)

For high rectal cancer (> 10 cm)

N: Number of patients in denominator who received neoadjuvant R(C)T

D: Number of patients in cStage I, treated with radical surgical resection with tumour in upper third

For mid rectal cancer ($>5 - 10$ cm)

N: Number of patients in denominator who received neoadjuvant R(C)T

D: Number of patients in cStage I, treated with radical surgical resection with tumour in middle third

For low rectal cancer (≤ 5 cm)

N: Number of patients in denominator who received neoadjuvant treatment

D: Number of patients in cStage I, treated with radical surgical resection with tumour in lower third

1.4.1.4 Proportion of cStage II-III patients treated with neoadjuvant 5-FU based chemoradiation, that received a continuous infusion of 5-FU (KCE 2008 QCI 1224; process indicator)

N: Number of patients in denominator that received a continuous infusion of 5-FU.

D: Number of patients with cStage II-III treated with radical surgical resection and long course pelvic chemoradiotherapy

Note Not used in PROCARE feedback until 2009 because not enough data. Solved retrospectively (at least partially by means of reminders in spring 2010). Also, alternative methods became available in the meantime (e.g. oral capecitabine).

1.4.1.5 *Proportion of cStage II-III patients treated with a long course of preoperative pelvic RT or chemoradiation, that completed this neoadjuvant treatment within the planned timing (KCE 2008 QCI 1225; process indicator)*

N: Number of patients in denominator for whom the radiotherapy treatment was not interrupted for more than five working days

D: Number of patients with cStage II-III who started with long course neoadjuvant radiotherapy which was followed by radical surgical resection

1.4.1.6 *Proportion of cStage II-III patients treated with a long course of preoperative pelvic RT or chemoradiation, that was operated 4 to 12 weeks after completion of the (chemo)radiation (KCE 2008 QCI 1226; process indicator)*

N: Number of patients in denominator that was operated 4 to 12 weeks after completion of the (chemo)radiotherapy

D: Number of patients with cStage II-III treated with long course neoadjuvant radiotherapy and for whom date of surgery and date of last irradiation are not missing

1.4.1.7 *Rate of acute grade 4 radio(chemo)therapy-related complications (KCE 2008 QCI 1227; process indicator)*

N: Number of patients in denominator that were presented acute grade 4 complications during/up to 8 weeks after completion of neoadjuvant or adjuvant (chemo)radiotherapy (long or short).

D: Number of patients treated with neoadjuvant or adjuvant radiotherapy and for whom follow-up data (at least until 1 year) are available.

Note Not used in PROCARE feedback until 2009 because not enough data. Solved retrospectively (at least partially by means of reminders in spring 2010).

1.5 SURGERY

Table 5: List of QCI in the domain of surgery

Code	Description	Type
1231	Proportion of R0 resections	Process
new	Distal margin involvement mentioned after SSO or Hartmann	Outcome
new	(y)p Distal margin involved (positive) after SSO or Hartmann for low rectal cancer (≤ 5 cm)	Outcome
new	Mesorectal (y)pCRM positivity after radical surgical resection	Outcome
1232a	Proportion of APR, Hartmann's procedure or total excision of colon and rectum with definitive ileostomy	Process
1232b	Proportion of patients with stoma 1 year after sphincter-sparing surgery	Outcome
new	Major leakage after PME + SSO + reconstruction	Outcome
new	Major leakage after TME + SSO + reconstruction (global, i.e. with or without primary derivative stoma)	Outcome
1234	Inpatient or 30-day mortality	Outcome
1235	Rate of intra-operative rectal perforation	Outcome
new	Postoperative major surgical morbidity with reintervention under narcosis after radical surgical resection	Outcome

1.5.1 Description of the QCIs

1.5.1.1 *Proportion of R0 resections (KCE 2008 QCI 1231; outcome indicator)*

Definitions:

- R0 status. Resections are classified as R0 if cM does not equal 'M1' and if type of resection at surgery is not 'R2' and if no one of the four criteria of R1 status are present.
- R1 status. Resections are classified as R1 if cM does not equal 'M1' and if type of resection at surgery is not 'R2' and if at least one of the following four conditions is present:
 - (y)pCRM < 1 mm
 - distal resection margin < 1 mm
 - rectum perforation as indicated by the surgeon
 - rectum perforation as indicated by the pathologist
- R2 status. Resections are classified as R2 if cM equals M1 and/or metastasis are discovered at surgery (and not completely resected).

Thus, if the type of resection at surgery is reported to be 'R2' then R status equals 'R2'.

- R status is reported as missing if cM status is missing and/or if data on two or more of the following criteria are missing: tumor free status of the (y)pCRM, the tumor free status of the distal resection margin, rectum perforation as indicated by the surgeon or pathologist.

R0 resection

N: Number of patients in denominator with R0 resection

D: Number of patients treated with radical surgical resection and for whom R status is not missing

R1 resection

N: Number of patients in denominator with R1 resection

D: Number of patients treated with radical surgical resection and for whom R status is not missing

R2 resection

N: Number of patients in denominator with R status equal 'R2'

D: Number of patients treated with radical surgical resection and for whom R status is not missing

1.5.1.2 Distal margin involvement mentioned after SSO or Hartmann (new QCI partially replacing KCE QCI 1231; outcome QCI)

N: Number of patients in denominator for whom it was reported whether the distal resection margin was invaded

D: Number of patients treated with Hartmann's procedure or SSO with reconstruction and for whom a pathology report sheet was completed

1.5.1.3 (y)p Distal margin involved (positive) after SSO or Hartmann for low rectal cancer (≤ 5 cm) (new QCI; outcome indicator)

N: Number of patients in denominator for whom the (y)p distal margin is invaded

D: Number of patients treated with Hartmann's procedure or SSO for rectal cancer in the lower third and for whom it is reported whether the (y)p distal margin is free or invaded

1.5.1.4 *Mesorectal (y)pCRM positivity after radical surgical resection (new QCI; outcome indicator)*

Note The definition of positivity (≤ 1 mm) differs with the definition of R1 status (invaded). It should apply only to the lateral margin of the mesorectum not to serosal positivity

Global

N: Number of patients in denominator for whom the mesorectal (y)pCRM ≤ 1 mm

D: Number of patients treated with radical surgical resection and for whom the mesorectal (y)pCRM is known

For high rectal cancer (> 10 cm)

N: Number of patients in denominator for whom the mesorectal (y)pCRM ≤ 1 mm

D: Number of patients treated with radical surgical resection with tumour in highest third and for whom (y)pCRM is known

For mid rectal cancer ($>5 - 10$ cm)

N: Number of patients in denominator for whom the mesorectal (y)pCRM ≤ 1 mm

D: Number of patients treated with radical surgical resection with tumour in middle third and for whom (y)pCRM is known

For low rectal cancer (≤ 5 cm)

N: Number of patients in denominator for whom the mesorectal (y)pCRM ≤ 1 mm

D: Number of patients treated with radical surgical resection with tumour in lowest third and for whom the mesorectal (y)pCRM is known

1.5.1.5 *Proportion of APR and Hartmann's procedure or total excision of colon and rectum with definitive ileostomy (KCE 2008 QCI 1232a; outcome indicator)*

Global (QCI)

N: Number of patients in denominator in whom APER or Hartmann's procedure or total excision of colon and rectum with definitive ileostomy was performed

D: Number of patients treated with any type of resection for rectal cancer at any known level

For high rectal cancer (> 10 cm)

N: Number of patients in denominator in whom APER or Hartmann's procedure or total excision of colon and rectum with definitive ileostomy was performed

D: Number of patients treated with any type of resection for tumour in upper third

For mid rectal cancer (>5 - 10 cm)

N: Number of patients in denominator in whom APER or Hartmann's procedure or total excision of colon and rectum with definitive ileostomy was performed

D: Number of patients treated with any type of resection for tumour in middle third

For low rectal cancer (≤ 5 cm)

N: Number of patients in denominator in whom APR or Hartmann's procedure or total excision of colon and rectum with definitive ileostomy was performed

D: Number of patients treated with any type of resection for tumour in lower third

1.5.1.6 *Proportion of patients with stoma 1 year after sphincter-sparing surgery (KCE 2008 QCI 1232b; outcome indicator)*

N: Number of patients in denominator still having a stoma 1 year after surgery

D: Number of patients treated with TME (complete rectum resection (TME) + straight CAA, coloplasty, pouch, side-to-end CAA, total excision of colon and rectum with IPAA, or another specified type of reconstruction) with a primary (constructed at the time of SSO) or secondary (constructed after SSO) derivative stoma or dismantling of anastomosis still alive 1 year after surgery and for whom follow-up at 1 year or more is known

1.5.1.7 *Rate of patients with major leakage of the anastomosis after PME + SSO + reconstruction (new QCI; outcome indicator)*

N: Number of patients with major leakage of the anastomosis (requiring reoperation for leakage)

D: Number of patients treated with PME (high or low anterior resection with colorectal anastomosis) and for whom it is reported whether there were postoperative complications or not

1.5.1.8 *Rate of patients with major leakage of the anastomosis after TME + SSO + reconstruction (global, i.e. with or without primary derivative stoma) (new QCI; outcome indicator)*

N: Number of patients with major leakage of the anastomosis (requiring reoperation for leakage)

D: Number of patients treated with TME (complete rectum resection (TME) + straight CAA, coloplasty, pouch, side-to-end CAA, total excision of colon and rectum with

IPAA, or another specified type of reconstruction) and for whom it is reported whether there were postoperative complications or not

1.5.1.9 *Inpatient or 30-day mortality (KCE 2008 QCI 1234; outcome indicator)*

N: Number of patients in denominator who died in hospital or within 30 days after surgery

D: Number of patients treated with radical surgical resection and for whom it is known whether they died in hospital or within 30 days after surgery and for whom the dates of surgery and survival or death are known.

1.5.1.10 *Rate of intra-operative rectal perforation (KCE 2008 QCI 1235; outcome indicator)*

N: Number of patients in denominator for whom the surgeon and/or pathologist reported rectal perforation

D: Number of patients treated with radical surgical resection and for whom perforation of the rectum (yes or no) is reported by either the surgeon or the pathologist

1.5.1.11 *Postoperative major surgical morbidity with reintervention under narcosis after radical surgical resection (new QCI; outcome indicator)*

N: Number of patients in denominator who presented major surgical morbidity requiring reintervention under narcosis

D: Number of patients treated with radical surgical resection and for whom postoperative data on morbidity/mortality are available

1.6 ADJUVANT TREATMENT

Table 6: List of QCI in the domain of adjuvant treatment

Code	Description	Type
1241	Proportion of (y)pStage III patients with R0 resection that received adjuvant chemotherapy within 3 months after surgery	Process
1242	Proportion of pStage II-III patients with R0 resection that received adjuvant radiotherapy or chemoradiotherapy within 3 months after surgery	Process
1243	Proportion of (y)pStage II-III patients with R0 resection that started adjuvant chemotherapy within 12 weeks after surgical resection	Process
1244	Proportion of (y)pStage II-III patients with R0 resection treated with adjuvant chemo(radio)therapy, that received 5-FU based chemotherapy	Process
1245	Rate of acute grade 4 chemotherapy-related complications	Process

1.6.1 Description of the QCIs

1.6.1.1 *Proportion of (y)pStage III patients with R0 resection that received adjuvant chemotherapy within 3 months after surgery (KCE 2008 QCI 1241; process indicator)*

N: Number of patients in denominator receiving adjuvant chemotherapy within 3 months after surgery

D: Number of patients treated with R0 radical surgical resection for (y)pStage III and for whom it is known whether they received adjuvant chemotherapy within 6 months after surgery or not.

1.6.1.2 *Proportion of pStage II-III patients with R0 resection that received adjuvant radiotherapy or chemoradiotherapy within 3 months after surgery (KCE 2008 QCI 1242; process indicator)*

N: Number of patients in denominator receiving adjuvant radio(chemo)therapy within 3 months after surgery

D: Number of patients treated with R0 radical surgical resection for pStage II or III without neoadjuvant treatment and for whom it is known whether they received adjuvant radio(chemo)therapy or not.

1.6.1.3 *Proportion of (y)pStage II-III patients with R0 resection that started adjuvant chemotherapy within 12 weeks after surgical resection (KCE 2008 QCI 1243; process indicator)*

N: Number of patients in denominator receiving adjuvant chemotherapy within 3 months after surgery

D: Number of patients treated with R0 radical surgical resection for (y)pStage II or III and for whom it is known whether they received adjuvant chemotherapy or not.

1.6.1.4 *Proportion of (y)pStage II-III patients with R0 resection treated with adjuvant chemo(radio)therapy, that received 5-FU based chemotherapy (KCE 2008 QCI 1244; process indicator)*

N: Number of patients in denominator receiving 5-fluorouracil based adjuvant chemotherapy

D: Number of patients who received adjuvant (radio)chemotherapy within 3 months after R0 radical surgical resection for (y)pStage II or III and for whom the type of adjuvant chemotherapy is known.

1.6.1.5 *Rate of acute grade 4 chemotherapy-related complications (KCE 2008 QCI 1245; process indicator)*

N: Number of patients in denominator that presented acute grade 4 complications during or within 4 weeks after completion of adjuvant chemo(radio)therapy

D: Number of patients treated with adjuvant chemotherapy and for whom follow-up data (at least until 1 year) are available.

Note Not used in PROCARE feedback until 2009 because not enough data. Solved retrospectively (at least partially by means of reminders in spring 2010).

1.7 PALLIATIVE TREATMENT

Table 7: List of QCI's in the domain of palliative treatment

Code	Description	Type
1251	Rate of cStage IV patients receiving chemotherapy	Process

1.7.1 Description of the QCI

1.7.1.1 *Rate of cStage IV patients receiving chemotherapy (KCE 2008 QCI 1251; process indicator)*

N: Number of patients in denominator that received chemotherapy

D: Number of patients with cStage IV and for whom it is known whether they received chemotherapy or not.

1.8 FOLLOW-UP

Table 8: List of QCIs in the domain of follow-up

Code	Description	Type
1261	Rate of curatively treated patients that received a colonoscopy within 1 year after resection	Process
1263	Late grade 4 complications of radiotherapy or chemoradiation	Outcome

1.8.1 Description of the QCIs

1.8.1.1 *Rate of curatively treated patients that received a colonoscopy within 1 year after resection (KCE 2008 QCI 1261; process indicator)*

N: Number of patients in denominator that received a colonoscopy

D: Number of patients treated with curative resection for c(p)Stage I-III and for whom follow-up data (at least until 2 years) are available.

Note Not used in PROCARE feedback until 2009 because not enough data. Solved retrospectively (at least partially by means of reminders in spring 2010).

1.8.1.2 *Late grade 4 complications of radiotherapy or chemoradiation (KCE 2008 QCI 1263; process indicator)*

N: Number of patients in denominator that presented late grade 4 complications after completion of (neo)adjuvant chemo(radio)therapy

D: Number of patients treated with neoadjuvant or adjuvant radio(chemo)therapy and for whom follow-up data (at least until 1 year) are available.

Note Not used in PROCARE feedback until 2009 because not enough data.

Grade refers to the severity of the AE. The CTCAE v3.0 displays Grades 1 through 5 with unique clinical descriptions of severity for each AE based on this general guideline:

- Grade 1 Mild AE
- Grade 2 Moderate AE
- Grade 3 Severe AE
- Grade 4 Life-threatening or disabling AE (An AE whose existence or immediate sequelae are associated with an imminent risk of death)
- Grade 5 Death related to AE

1.9 HISTOPATHOLOGIC EXAMINATION

Table 9: List of QCIs in the domain of histopathologic examination

Code	Description	Type
1271	Use of the pathology report sheet	Process
1272	Quality of TME assessed according to Quirke and mentioned in the pathology report	Process
1273	Distal tumour-free margin mentioned in the pathology report	Process
1274	Number of lymph nodes examined	Process
1275	(y)pCRM mentioned in mm in the pathology report	Process
1276	Tumour regression grade mentioned in the pathology report (after neoadjuvant treatment)	Process

1.9.1 Description of the QCIs

1.9.1.1 *Use of the pathology report sheet (KCE 2008 QCI 1271; process indicator)*

N: Number of patients in denominator for whom a pathology report sheet was completed

D: Number of patients treated with (local or radical) resection and for whom date of resection is later than or equal to the 1st of January 2007.

1.9.1.2 *Quality of TME assessed according to Quirke and mentioned in the pathology report (KCE 2008 QCI 1272; process indicator)*

N: Number of patients for whom the external surface of TME was reported in the pathology report sheet

D: Number of patients treated with TME as indicated by the surgeon after the 1st of January 2007.

1.9.1.3 *Distal tumour-free margin mentioned in the pathology report (KCE 2008 QCI 1273; process indicator)*

N: Number of patients in denominator for whom the length of the distal free tumour free margin was reported in the pathology report

D: Number of patients treated with SSO or Hartmann's procedure.

1.9.1.4 *Number of lymph nodes examined (KCE 2008 QCI 1274; process indicator)*

The median number of lymph nodes examined is computed for the following conditions:

- no or short course neoadjuvant RT

- long course neoadjuvant RT
- course type missing

1.9.1.5 *(y)pCRM mentioned in mm in the pathology report (KCE 2008 QCI 1275; process indicator)*

N: Number of patients in denominator for whom the mesorectal (y)pCRM was mentioned in the pathology report

D: Number of patients treated with radical surgical resection and for whom a pathology report was completed

1.9.1.6 *Tumour regression grade mentioned in the pathology report (after neoadjuvant treatment) (KCE 2008 QCI 1276; process indicator)*

N: Number of patients in denominator having their tumour regression grade mentioned in the pathology report

D: Number of patients treated with neoadjuvant long course radio(chemo)therapy and surgery

2 OUTCOME-SPECIFIC QUALITY OF CARE INDICATORS

Table 10: List of outcome-specific QCI's over all domains

Code	Description
1111	Overall 5-year survival by stage
1112	Disease-specific 5-year survival by stage
new	Relative survival
1113	Proportion of patients with local recurrence
new	Disease-free survival
new	Distal margin involvement mentioned after SSO or Hartmann
new	(y)p Distal margin involved (positive) after SSO or Hartmann for low rectal cancer (≤ 5 cm)
new	Mesorectal (y)pCRM positivity after radical surgical resection
1232b	Proportion of patients with stoma 1 year after sphincter-sparing surgery
new	Major leakage after PME + SSO + reconstruction
new	Major leakage after TME + SSO + reconstruction (global, i.e. with or without primary derivative stoma)
1234	Inpatient or 30-day mortality
1235	Rate of intra-operative rectal perforation
new	Postoperative major surgical morbidity with reintervention under narcosis after radical surgical resection
1263	Late grade 4 complications of radiotherapy or chemoradiation

3 PROCESS-SPECIFIC QUALITY OF CARE INDICATORS

Table 11: List of process-specific QCI's over all domains

Code	Description
1211	Proportion of patients with a documented distance from the anal verge
1212	Proportion of patients in whom a CT of the abdomen and RX or CT thorax was performed before any treatment
1213	Proportion of patients in whom a CEA was performed before any treatment
1214	Proportion of patients undergoing elective surgery that had preoperative complete large bowel-imaging
1215	Proportion of patients in whom a TRUS and pelvic CT and/or pelvic MRI was performed before any treatment
1216	Proportion of patients with cStage II-III RC that have a reported cCRM
1217	Time between first histopathologic diagnosis and first treatment
new	Accuracy of cM0 staging
new	Accuracy of cT/cN staging if no or short radiotherapy (separately presented in 2 tables)
new	Use of TRUS in cT1/cT2
new	Use of MRI in cStage II or III
new	Proportion of cStage II-III patients that received a neoadjuvant pelvic RT
new	Proportion of patients with cCRM \leq 2 mm on MRI/CT that received long course neoadjuvant radio(chemo)therapy
new	Proportion of patients with cStage I that received neoadjuvant radio(chemo)therapy
1224	Proportion of cStage II-III patients treated with neoadjuvant 5-FU based chemoradiation, that received a continuous infusion of 5-FU
1225	Proportion of cStage II-III patients treated with a long course of preoperative pelvic RT or chemoradiation, that completed this neoadjuvant treatment within the planned timing
1226	Proportion of cStage II-III patients treated with a long course of preoperative pelvic RT or chemoradiation, that was operated 4 to 12 weeks after completion of the (chemo)radiation
1227	Rate of acute grade 4 radio(chemo)therapy-related complications
1231	Proportion of R0 resections
1232a	Proportion of APR, Hartmann's procedure or total excision of colon and rectum with definitive ileostomy
1241	Proportion of (y)pStage III patients with R0 resection that received adjuvant chemotherapy within 3 months after surgery
1242	Proportion of pStage II-III patients with R0 resection that received adjuvant radiotherapy or chemoradiotherapy within 3 months after surgery
1243	Proportion of (y)pStage II-III patients with R0 resection that started adjuvant chemotherapy within 12 weeks after surgical resection
1244	Proportion of (y)pStage II-III patients with R0 resection treated with adjuvant chemo(radio)therapy, that received 5-FU based chemotherapy
1245	Rate of acute grade 4 chemotherapy-related complications

1251	Rate of cStage IV patients receiving chemotherapy
1261	Rate of curatively treated patients that received a colonoscopy within 1 year after resection
1271	Use of the pathology report sheet
1272	Quality of TME assessed according to Quirke and mentioned in the pathology report
1273	Distal tumour-free margin mentioned in the pathology report
1274	Number of lymph nodes examined
1275	(y)pCRM mentioned in mm in the pathology report
1276	Tumour regression grade mentioned in the pathology report (after neoadjuvant treatment)

This page is left intentionally blank.

KCE reports

- 33 Effects and costs of pneumococcal conjugate vaccination of Belgian children. D/2006/10.273/54.
- 34 Trastuzumab in Early Stage Breast Cancer. D/2006/10.273/25.
- 36 Pharmacological and surgical treatment of obesity. Residential care for severely obese children in Belgium. D/2006/10.273/30.
- 37 Magnetic Resonance Imaging. D/2006/10.273/34.
- 38 Cervical Cancer Screening and Human Papillomavirus (HPV) Testing D/2006/10.273/37.
- 40 Functional status of the patient: a potential tool for the reimbursement of physiotherapy in Belgium? D/2006/10.273/53.
- 47 Medication use in rest and nursing homes in Belgium. D/2006/10.273/70.
- 48 Chronic low back pain. D/2006/10.273/71.
- 49 Antiviral agents in seasonal and pandemic influenza. Literature study and development of practice guidelines. D/2006/10.273/67.
- 54 Cost-effectiveness analysis of rotavirus vaccination of Belgian infants D/2007/10.273/11.
- 59 Laboratory tests in general practice D/2007/10.273/26.
- 60 Pulmonary Function Tests in Adults D/2007/10.273/29.
- 64 HPV Vaccination for the Prevention of Cervical Cancer in Belgium: Health Technology Assessment. D/2007/10.273/43.
- 65 Organisation and financing of genetic testing in Belgium. D/2007/10.273/46.
- 66 Health Technology Assessment: Drug-Eluting Stents in Belgium. D/2007/10.273/49.
- 70. Comparative study of hospital accreditation programs in Europe. D/2008/10.273/03
- 71. Guidance for the use of ophthalmic tests in clinical practice. D/2008/10.273/06.
- 72. Physician workforce supply in Belgium. Current situation and challenges. D/2008/10.273/09.
- 74 Hyperbaric Oxygen Therapy: a Rapid Assessment. D/2008/10.273/15.
- 76. Quality improvement in general practice in Belgium: status quo or quo vadis? D/2008/10.273/20
- 82. 64-Slice computed tomography imaging of coronary arteries in patients suspected for coronary artery disease. D/2008/10.273/42
- 83. International comparison of reimbursement principles and legal aspects of plastic surgery. D/2008/10.273/45
- 87. Consumption of physiotherapy and physical and rehabilitation medicine in Belgium. D/2008/10.273/56
- 90. Making general practice attractive: encouraging GP attraction and retention D/2008/10.273/66.
- 91 Hearing aids in Belgium: health technology assessment. D/2008/10.273/69.
- 92. Nosocomial Infections in Belgium, part I: national prevalence study. D/2008/10.273/72.
- 93. Detection of adverse events in administrative databases. D/2008/10.273/75.
- 95. Percutaneous heart valve implantation in congenital and degenerative valve disease. A rapid Health Technology Assessment. D/2008/10.273/81
- 100. Threshold values for cost-effectiveness in health care. D/2008/10.273/96
- 102. Nosocomial Infections in Belgium: Part II, Impact on Mortality and Costs. D/2009/10.273/03
- 103 Mental health care reforms: evaluation research of 'therapeutic projects' - first intermediate report. D/2009/10.273/06.
- 104. Robot-assisted surgery: health technology assessment. D/2009/10.273/09
- 108. Tiotropium in the Treatment of Chronic Obstructive Pulmonary Disease: Health Technology Assessment. D/2009/10.273/20
- 109. The value of EEG and evoked potentials in clinical practice. D/2009/10.273/23
- 111. Pharmaceutical and non-pharmaceutical interventions for Alzheimer's Disease, a rapid assessment. D/2009/10.273/29
- 112. Policies for Orphan Diseases and Orphan Drugs. D/2009/10.273/32.
- 113. The volume of surgical interventions and its impact on the outcome: feasibility study based on Belgian data
- 114. Endobronchial valves in the treatment of severe pulmonary emphysema. A rapid Health Technology Assessment. D/2009/10.273/39
- 115. Organisation of palliative care in Belgium. D/2009/10.273/42
- 116. Interspinous implants and pedicle screws for dynamic stabilization of lumbar spine: Rapid assessment. D/2009/10.273/46

117. Use of point-of care devices in patients with oral anticoagulation: a Health Technology Assessment. D/2009/10.273/49.
118. Advantages, disadvantages and feasibility of the introduction of 'Pay for Quality' programmes in Belgium. D/2009/10.273/52.
119. Non-specific neck pain: diagnosis and treatment. D/2009/10.273/56.
121. Feasibility study of the introduction of an all-inclusive case-based hospital financing system in Belgium. D/2010/10.273/03
122. Financing of home nursing in Belgium. D/2010/10.273/07
123. Mental health care reforms: evaluation research of 'therapeutic projects' - second intermediate report. D/2010/10.273/10
124. Organisation and financing of chronic dialysis in Belgium. D/2010/10.273/13
125. Impact of academic detailing on primary care physicians. D/2010/10.273/16
126. The reference price system and socioeconomic differences in the use of low cost drugs. D/2010/10.273/20.
127. Cost-effectiveness of antiviral treatment of chronic hepatitis B in Belgium. Part I: Literature review and results of a national study. D/2010/10.273/24.
128. A first step towards measuring the performance of the Belgian healthcare system. D/2010/10.273/27.
129. Breast cancer screening with mammography for women in the agegroup of 40-49 years. D/2010/10.273/30.
130. Quality criteria for training settings in postgraduate medical education. D/2010/10.273/35.
131. Seamless care with regard to medications between hospital and home. D/2010/10.273/39.
132. Is neonatal screening for cystic fibrosis recommended in Belgium? D/2010/10.273/43.
133. Optimisation of the operational processes of the Special Solidarity Fund. D/2010/10.273/46.
135. Emergency psychiatric care for children and adolescents. D/2010/10.273/51.
136. Remote monitoring for patients with implanted defibrillator. Technology evaluation and broader regulatory framework. D/2010/10.273/55.
137. Pacemaker therapy for bradycardia in Belgium. D/2010/10.273/58.
138. The Belgian health system in 2010. D/2010/10.273/61.
139. Guideline relative to low risk birth. D/2010/10.273/64.
140. Cardiac rehabilitation: clinical effectiveness and utilisation in Belgium. d/2010/10.273/67.
141. Statins in Belgium: utilization trends and impact of reimbursement policies. D/2010/10.273/71.
142. Quality of care in oncology: Testicular cancer guidelines. D/2010/10.273/74
143. Quality of care in oncology: Breast cancer guidelines. D/2010/10.273/77.
144. Organization of mental health care for persons with severe and persistent mental illness. What is the evidence? D/2010/10.273/80.
145. Cardiac resynchronisation therapy. A Health technology Assessment. D/2010/10.273/84.
146. Mental health care reforms: evaluation research of 'therapeutic projects'. D/2010/10.273/87
147. Drug reimbursement systems: international comparison and policy recommendations. D/2010/10.273/90
149. Quality indicators in oncology: testis cancer. D/2010/10.273/98.
150. Quality indicators in oncology: breast cancer. D/2010/10.273/101.
153. Acupuncture: State of affairs in Belgium. D/2011/10.273/06.
154. Homeopathy: State of affairs in Belgium. D/2011/10.273/14.
155. Cost-effectiveness of 10- and 13-valent pneumococcal conjugate vaccines in childhood. D/2011/10.273/21.
156. Home Oxygen Therapy. D/2011/10.273/25.
158. The pre-market clinical evaluation of innovative high-risk medical devices. D/2011/10.273/31
159. Pharmacological prevention of fragility fractures in Belgium. D/2011/10.273/34.
160. Dementia: which non-pharmacological interventions? D/2011/10.273/37
161. Quality Insurance of rectal cancer – phase 3: statistical methods to benchmark centers on a set of quality indicators. D/2011/10.273/40.

This list only includes those KCE reports for which a full English version is available. However, all KCE reports are available with a French or Dutch executive summary and often contain a scientific summary in English.

