

Strengthening the Reporting of Genetic Risk Prediction Studies: The GRIPS Statement

A. CECILE J.W. JANSSENS⁽¹⁾, JOHN P.A. IOANNIDIS^(2,3,4,5,6), CORNELIA M. VAN DUIJN⁽¹⁾, JULIAN LITTLE⁽⁷⁾, MUIN J. KHOURY⁽⁸⁾, FOR THE GRIPS GROUP

ABSTRACT

The rapid and continuing progress in gene discovery for complex diseases is fuelling interest in the potential application of genetic risk models for clinical and public health practice.

The number of studies assessing the predictive ability is steadily increasing, but the quality and completeness of reporting varies.

A multidisciplinary workshop sponsored by the Human Genome Epidemiology Network developed a checklist of 25 items recommended for strengthening the reporting of Genetic Risk Prediction Studies (GRIPS), building on the principles established by prior reporting guidelines.

These recommendations aim to enhance the transparency of study reporting, and thereby to improve the synthesis and application of information from multiple studies that might differ in design, conduct or analysis.

A detailed Explanation and Elaboration document is published separately in PLoS Medicine as Supporting Information (S1) and in Janssens AC, Ioannidis JP, van Duijn CM, Little J, Khoury MJ; GRIPS Group. Strengthening the reporting of genetic risk prediction studies: the GRIPS statement. *Eur J Epidemiol* 2011; 26(4): 255-9.

Key words: Genetic risk models, GRIPS

(1) Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands.

(2) Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece.

(3) Biomedical Research Institute, Foundation for Research and Technology, Ioannina, Greece.

(4) Department of Medicine, Tufts University School of Medicine, Boston MA, USA.

(5) Center for Genetic Epidemiology and Modeling and Tufts CTSI, Institute for Clinical Research and Health Policy Studies, Tufts Medical Center, Boston MA, USA.

(6) Department of Epidemiology, Harvard School of Public Health, Boston MA, USA.

(7) Department of Epidemiology and Community Medicine, University of Ottawa, Ottawa ON, Canada.

(8) Office of Public Health Genomics, Centers for Disease Control and Prevention, Atlanta GA, USA.

CORRESPONDING AUTHOR: A. Cecile J.W. Janssens, Erasmus University Medical Center, Department of Epidemiology, PO Box 2040, 3000 CA Rotterdam, the Netherlands. Email: a.janssens@erasmusmc.nl; Telephone: +31-10-7044214; Fax: +31-10-7044657

INTRODUCTION

The recent successes of genome-wide association studies and the promises of whole genome sequencing fuel the interest in the

translation of this new wave of basic genetic knowledge to health care practice. Knowledge about genetic risk factors may be used to target diagnostic, preventive and therapeutic interventions for complex disorders based on a

person's genetic risk, or to complement existing risk models based on classical non-genetic factors such as the Framingham risk score for cardiovascular disease. Implementation of genetic risk prediction in health care requires a series of studies that encompass all phases of translational research (1, 2), starting with a comprehensive evaluation of genetic risk prediction.

With increasing numbers of discovered genetic markers that can be used in future genetic risk prediction studies, it is crucial to enhance the quality of the reporting of these studies, since valid interpretation could be compromised by the lack of reporting of key information. Information that is often missing includes details in the description of how the study was designed and conducted (e.g., how genetic variants were selected and coded, how risk models or genetic risk scores were constructed and how risk categories were chosen), or how the results should be interpreted. An appropriate assessment of the study's strengths and weaknesses is not possible without this information. There is ample evidence that prediction research often suffers from poor design and biases, and these may have an impact also on the results of the studies and on models of disease outcomes based on these studies (3-5). Although most prognostic studies published to date claim significant results (6, 7), very few translate to clinically useful applications. Just as for observational epidemiological studies (8), poor reporting complicates the use of the specific study for research, clinical, or public health purposes and hampers the synthesis of evidence across studies.

Reporting guidelines have been published for various research designs (9) and these contain many items that are also relevant to genetic risk prediction studies. In particular, the guidelines for genetic association studies (STREGA) have relevant items on the assessment of genetic variants, and the guidelines for observational studies (STROBE) have relevant items about the reporting of study design. The guidelines for diagnostic studies (STARD) and those for tumor marker prognostic studies (REMARK) include relevant items about test evaluation, and the REMARK guidelines relevant items about risk prediction (10-13). However, none of these guidelines are fully suited to genetic risk prediction studies, an emerging field of investigations with specific methodological issues that need to be addressed, such as the handling of large numbers of genetic variants (from 10s to 10

000s), which come with greater challenges and flexibility on how these can be dealt with in the analyses. We organized a two day workshop with an international group of risk prediction researchers, epidemiologists, geneticists, methodologists, statisticians and journal editors to develop recommendations for the reporting of Genetic Risk Prediction Studies (GRIPS).

Genetic risk prediction studies

Genetic risk prediction studies typically develop or validate models for the risk of disease, but they are also being investigated for prognostic outcome, treatment response or treatment-related harms. Risk prediction models are statistical algorithms, which may be simple genetic risk scores (e.g., risk allele counts), be based on regression analyses (e.g., weighted risk scores or predicted risks) or on more complex analytic approaches such as support vector machine learning or classification trees. The risk models may be based on genetic variants only, or include both genetic and non-genetic risk factors (14).

Aims and use of the GRIPS Statement

The 25 items of the GRIPS Statement are intended to maximize the transparency, quality and completeness of reporting on research methodology and findings in a particular study. It is important to emphasize that these recommendations are guidelines only for how to report research, and do not prescribe how to perform genetic risk prediction studies. The guidelines do not support or oppose the choice of any particular study design or method, e.g., the guidelines recommend that the study population should be described, but do not specify which population is preferred in a particular study.

The intended audience for the reporting guidelines is broad and includes epidemiologists, geneticists, statisticians, clinician scientists and laboratory-based investigators who undertake genetic risk prediction studies, as well as journal editors and reviewers who have to appraise the design, conduct and analysis of such studies. In addition, it includes 'users' of such studies who wish to understand the basic premise, design, and limitations of genetic prediction studies in order to interpret the results for their potential application

in health care. These guidelines are also intended to ensure that essential data from future genetic risk prediction studies are presented, which will facilitate information synthesis as part of systematic reviews and meta-analyses.

Items presented in the checklist are relevant for a wide array of risk prediction studies, because GRIPS focuses on the main aspects of the design and analysis of risk prediction studies. GRIPS does not address randomized trials that may be performed to test risk models, nor does it specifically address decision analyses, cost-effectiveness analyses, assessment of health care needs or assessment of barriers to health care implementation (15). Once the performance of a risk model has been established, these next steps towards implementation require further evaluation (10, 16). For the reporting of these studies, which go beyond the assessment of genetic risk models as such, additional requirements apply. However, proper documentation of genetic predictive research according to GRIPS might facilitate the translation of research findings into clinical and public health practice.

Development of the GRIPS Statement

The GRIPS Statement were developed by a multidisciplinary panel of 25 risk prediction researchers, epidemiologists, geneticists, methodologists, statisticians and journal editors, seven of whom were also part of the STREGA Initiative (11). They attended a two-day meeting in Atlanta, GA, USA, in December 2009 sponsored by the Centers for Disease Control and Prevention on behalf of the Human Genome Epidemiology Network (HuGENet) (17). Participants discussed a draft version of the guidelines that was prepared and distributed prior to the meeting. This draft version was developed based on existing reporting guidelines, namely STREGA (11), REMARK (13), and STARD (12). These were selected out of all available guidelines (see www.equator-network.org) because of their focus on observational study designs and genetic factors (STREGA), prediction models (REMARK), and test evaluation (REMARK and STARD). During the meeting, methodological issues pertinent to risk prediction studies were addressed in presentations. Workshop participants were asked to change, combine or delete proposed items and add additional items if necessary. Participants had extensive post-meeting electronic correspondence. To harmonize our recommendations for genetic risk

prediction studies with previous guidelines, we chose the same wording for the items wherever possible. Finally, we tried to have consistency with previous guidelines for the evaluation of risk prediction studies of cardiovascular diseases and cancer (2, 18). The final version of the checklist is presented in Table 1.

The GRIPS Explanation and Elaboration paper

Accompanying this GRIPS statement, an Explanation and Elaboration paper has been written (see S1 in PLoS Medicine), modeled after those developed for other reporting guidelines (19-22). The Explanation and Elaboration paper illustrates each item with at least one published example that we consider being transparent in reporting, explains the rationale for its inclusion in the checklist, and presents details of the item that need to be addressed to ensure transparent reporting. The Explanation and Elaboration paper was produced after the meeting. The document was prepared by a small subgroup and shared with all workshop participants for additional revisions and final approval.

Concluding remarks and future directions

High quality reporting reveals the strengths and weaknesses of empirical studies, facilitates the interpretation of the scientific and health care relevance of the results, in particular within the framework of systematic reviews and meta-analyses, and helps build a solid evidence base for moving genomic discoveries into applications in health care practice. The GRIPS guidelines were developed to improve the transparency, quality and completeness of the reporting of genetic risk prediction studies. As outlined in the introduction, GRIPS does not prescribe how studies should be designed, conducted and analyzed, and therefore, the guidelines should not be used to assess the quality of empirical studies (23). The guidelines should only be used to check whether all essential items are adequately reported.

Finally, the methodology for designing and assessing genetic risk prediction models is still developing. For example, newer measures of reclassification were first introduced in 2007 (24) and several alternative reclassification measures have been proposed (25). Which measures to apply and when to use measures of reclassification is still subject to ongoing evaluation and discussion (26).

TABLE 1

| REPORTING RECOMMENDATIONS FOR EVALUATIONS OF RISK PREDICTION MODELS THAT INCLUDE GENETIC VARIANTS | | |
|---|----|--|
| TITLE & ABSTRACT | | |
| | 1 | (a) Identify the article as a study of risk prediction using genetic factors. (b) Use recommended keywords in the abstract: genetic or genomic, risk, prediction. |
| INTRODUCTION | | |
| Background and Rationale | 2 | Explain the scientific background and rationale for the prediction study. |
| Objectives | 3 | Specify the study objectives and state the specific model(s) that is/are investigated. State if the study concerns the development of the model(s), a validation effort, or both. |
| METHODS | | |
| Study design and setting | 4* | Specify the key elements of the study design and describe the setting, locations and relevant dates, including periods of recruitment, follow-up and data collection. |
| Participants | 5* | Describe eligibility criteria for participants, and sources and methods of selection of participants. |
| Variables: definition | 6* | Clearly define all participant characteristics, risk factors and outcomes. Clearly define genetic variants using a widely-used nomenclature system. |
| Variables: assessment | 7* | (a) Describe sources of data and details of methods of assessment (measurement) for each variable. (b) Give a detailed description of genotyping and other laboratory methods. |
| Variables: coding | 8 | (a) Describe how genetic variants were handled in the analyses. (b) Explain how other quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why. |
| Analysis: risk model construction | 9 | Specify the procedure and data used for the derivation of the risk model. Specify which candidate variables were initially examined or considered for inclusion in models. Include details of any variable selection procedures and other model-building issues. Specify the horizon of risk prediction (e.g., 5-year risk). |
| Analysis: validation | 10 | Specify the procedure and data used for the validation of the risk model. |
| Analysis: missing data | 11 | Specify how missing data were handled. |
| Analysis: statistical methods | 12 | Specify all measures used for the evaluation of the risk model including, but not limited to, measures of model fit and predictive ability. |
| Analysis: other | 13 | Describe all subgroups, interactions and exploratory analyses that were examined. |

* Marked items should be reported for every population in the study

Furthermore, alternative strategies for constructing risk models other than simple regression analyses are being explored, and these may add increasing complexity to the reporting. In formulating the

items of the GRIPS Statement, these methodological advances were anticipated. It was for this reason that the GRIPS Statement recommends how a study should be reported and not how a study should be

TABLE 1 (CONTINUED)

| REPORTING RECOMMENDATIONS FOR EVALUATIONS OF RISK PREDICTION MODELS THAT INCLUDE GENETIC VARIANTS | | |
|---|-----|---|
| RESULTS | | |
| Participants | 14* | Report the numbers of individuals at each stage of the study. Give reasons for non-participation at each stage. Report the number of participants not genotyped, and reasons why they were not genotyped. |
| Descriptives: population | 15* | Report demographic and clinical characteristics of the study population, including risk factors used in the risk modeling. |
| Descriptives: model estimates | 16 | Report unadjusted associations between the variables in the risk model(s) and the outcome. Report adjusted estimates and their precision from the full risk model(s) for each variable. |
| Risk distributions | 17* | Report distributions of predicted risks and/or risk scores. |
| Assessment | 18 | Report measures of model fit and predictive ability, and any other performance measures, if pertinent. |
| Validation | 19 | Report any validation of the risk model(s). |
| Other analyses | 20 | Present results of any subgroup, interaction or exploratory analyses, whenever pertinent. |
| DISCUSSION | | |
| Limitations | 21 | Discuss limitations and assumptions of the study, particularly those concerning study design, selection of participants, measurements and analyses, and discuss their impact on the results of the study. |
| Interpretation | 22 | Give an overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence. |
| Generalizability | 23 | Discuss the generalizability and, if pertinent, the health care relevance of the study results. |
| OTHER | | |
| Supplementary information | 24 | State whether databases for the analyzed data, risk models and/or protocols are or will become publicly available and if so, how they can be accessed. |
| Funding | 25 | Give the source of funding and the role of the funders for the present study. State whether there are any conflicts of interest. |

* Marked items should be reported for every population in the study

conducted or analyzed. Therefore, methodological and analytical developments will not immediately impact the validity and relevance of the items, but the GRIPS statement will be updated when this is warranted by essential new developments in the construction and evaluation of genetic risk models.

ACKNOWLEDGEMENTS: the following individuals have contributed to the content and elaboration of the GRIPS Statement: A. Cecile J.W. Janssens, John P.A. Ioannidis,

Sara Bedrosian, Paolo Boffetta, Siobhan M. Dolan, Nicole Dowling, Isabel Fortier, Andrew N. Freedman, Jeremy M. Gimsbaw, Jeffrey Gulcher, Marta Gwinn, Mark A. Hlatky, Holly James, Peter Kraft, Stephanie Melillo, Christopher J. O'Donnell, Michael J. Pencina, David Ransohoff, Sheri D Schully, Daniela Seminara, Deborah M. Winn, Caroline F. Wright, Cornelia M. van Duijn, Julian Little, Muin J. Khoury. The findings and conclusions in this report are those of the authors and do not necessarily reflect the views of the Department of Health and Human Services.

AUTHOR CONTRIBUTIONS: the authors coordinated the GRIPS Initiative and contributed to the writing of the paper. ACJWJ and MJK initiated the GRIPS Initiative, ACJWJ wrote the first draft of the guidelines and the paper and MJK was responsible for the coordination of the meeting.

FUNDING: A. Cecile J.W. Janssens is financially supported by grants from the Erasmus University Medical Center Rotterdam, the Center for Medical Systems Biology in the framework of the Netherlands Genomics Initiative (NGI)

and the VIDI grant of the Netherlands Organisation for Scientific Research (NWO).

John P.A. Ioannidis: Tufts CTSI is supported by the National Institutes of Health/National Center for Research Resources (UL1 RR025752). Opinions in this paper are those of the authors and do not necessarily represent the official position or policies of the Tufts CTSI.

Julian Little holds a Canada Research Chair in Human Genome Epidemiology.

COMPETING INTERESTS: the authors are not aware of any other potential conflict of interest associated with this manuscript.

ABBREVIATIONS: GRIPS: Genetic Risk Prediction Studies, HuGENet: Human Genome Epidemiology Network; REMARK: Guidelines for Reporting of tumor MARKer studies; STARD: STAndards for Reporting Diagnostic accuracy; STREGA: STrengthening the REporting of Genetic Association studies; STROBE: STrengthening the Reporting of OBServational studies in Epidemiology

References

- (1) Khoury MJ, Gwinn M, Yoon PW, et al. The continuum of translation research in genomic medicine: how can we accelerate the appropriate integration of human genome discoveries into health care and disease prevention? *Genet Med* 2007; 9: 665-74
- (2) Hlatky MA, Greenland P, Arnett DK, et al. Criteria for evaluation of novel markers of cardiovascular risk: a scientific statement from the American Heart Association. *Circulation* 2009; 119: 2408-16
- (3) Kyzas PA, Denaxa-Kyza D, Ioannidis JP. Quality of reporting of cancer prognostic marker studies: association with reported prognostic effect. *J Natl Cancer Inst* 2007; 99: 236-43
- (4) Kyzas PA, Loizou KT, Ioannidis JP. Selective reporting biases in cancer prognostic factor studies. *J Natl Cancer Inst* 2005; 97: 1043-55
- (5) McShane LM, Altman DG, Sauerbrei W, et al. REporting recommendations for tumor MARKer prognostic studies (REMARK). *Nat Clin Pract Urol* 2005; 2: 416-22
- (6) Kyzas PA, Denaxa-Kyza D, Ioannidis JP. Almost all articles on cancer prognostic markers report statistically significant results. *Eur J Cancer* 2007; 43: 2559-79
- (7) Tzoulaki I, Liberopoulos G, Ioannidis JP. Assessment of claims of improved prediction beyond the Framingham risk score. *JAMA* 2009; 302: 2345-52
- (8) von Elm E, Egger M. The scandal of poor epidemiological research. *BMJ* 2004; 329: 868-9
- (9) Simera I, Moher D, Hoey J, Schulz KF, Altman DG. A catalogue of reporting guidelines for health research. *Eur J Clin Invest* 2010; 40: 35-53
- (10) von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med* 2007; 4: e296
- (11) Little J, Higgins JP, Ioannidis JP, et al. STrengthening the REporting of Genetic Association Studies (STREGA): an extension of the STROBE statement. *Ital J Public Health* 2009; 6: 238-55
- (12) Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. *BMJ* 2003; 326: 41-4
- (13) McShane LM, Altman DG, Sauerbrei W, et al. Reporting recommendations for tumor marker prognostic studies. *J Clin Oncol* 2005; 23: 9067-72
- (14) Janssens ACJW, Van Duijn CM. Genome-based prediction of common diseases: methodological considerations for future research. *Genome Med* 2009; 1: 20
- (15) Khoury MJ, Gwinn M, Ioannidis JP. The emergence of translational epidemiology: from scientific discovery to population health impact. *Am J Epidemiol* 2010; 172: 517-24
- (16) Moons KG, Altman DG, Vergouwe Y, Royston P. Prognosis and prognostic research: application and impact of prognostic models in clinical practice. *BMJ* 2009; 338: b606
- (17) Khoury MJ, Dorman JS. The Human Genome Epidemiology Network. *Am J Epidemiol* 1998; 148: 1-3

- (18) Freedman AN, Seminara D, Gail MH, Hartge P, Colditz GA et al. Cancer risk prediction models: a workshop on development, evaluation, and application. *J Natl Cancer Inst* 2005; 97: 715-23
- (19) Altman DG, Schulz KF, Moher D, et al. The revised CONSORT statement for reporting randomized trials: explanation and elaboration. *Ann Intern Med* 2001; 134: 663-94
- (20) Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med* 2003; 138: W1-12
- (21) Liberati A, Altman DG, Tetzlaff J, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ital J Public Health* 2009; 6: 354-91
- (22) Vandembroucke JP, von Elm E, Altman DG, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med* 2007; 4: e297
- (23) Vandembroucke JP. STREGA, STROBE, STARD, SQUIRE, MOOSE, PRISMA, GNOSIS, TREND, ORION, COREQ, QUOROM, REMARK... and CONSORT: for whom does the guideline toll? *J Clin Epidemiol* 2009; 62: 594-6
- (24) Cook NR. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 2007; 115: 928-35
- (25) Pencina MJ, D'Agostino RB, Sr., D'Agostino RB, Jr., Vasan RS. Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond. *Stat Med* 2008; 27: 157-72
- (26) Janssens ACJW, Khoury MJ. Assessment of improved prediction beyond traditional risk factors: when does a difference make a difference? *Circ Cardiovasc Genet* 2010; 3: 3-5

