FINAL PROJECT REPORT
Canola Agronomic Research Program (CARP)

The Annual Final Report should fully describe the work completed for the year and note the personnel involved. It should also note any deviations from the original plan and next and/or corrective steps as may be required if deviations are noted. The report should also provide an update on the status of the Project including forecasted date of completion. A complete statement of expenses should be included. In the event major changes are anticipated within the budget supporting notes along with a proposed budget should also be included. The report should also capture a complete summary of activity for the year.

Project Title: CARP Project No. 2011-09 - Advanced statistical analysis of small-plot canola variety trial data

Research Team Information

Lead Researchers:		
Name	**Institution**	**Expertise Added**
Dr. Anita Brûlé-Babel	Dept. of Plant Science University of Manitoba	Many years as Professor of Plant Breeding and Genetics
Research Team Members		
Name	**Institution**	**Expertise Added**
Dr. Gary Crow	Dept. of Animal Science University of Manitoba	Many years as Professor of Animal Genetics and Breeding, and Biometrician (statistics); retired 2012
Lyle Friesen	Dept. of Plant Science University of Manitoba	Research Associate. Statistical analysis, particularly Mixed Model analysis of large crop variety datasets using the specialized statistical computer program, ASReml.

Project Start Date: April 1, 2011 **Project Completion Date:** September 30, 2014

Reporting Period: Final Project Report

CARP Project Number: 2011-09

Instructions: This Annual Report shall be completed and submitted on or about March 31st of each fiscal year that the agreement is in effect. The Lead Researcher of the project in question shall complete and submit the report on behalf of his/her complete research team.

This Report is a means by which to provide a detailed update on the status of the project and summarize project activities. Details may be general in nature unless major issues or changes arise (e.g., change of scientists, significant change or delay of activities) including impacts on budgets. Please note that financial reports of major impact on budgets.

The following template is provided to assist you in completing this task. Please forward the completed document electronically to your appropriate CCC contact.

1. Forecasted Date of Completion:

September 30, 2014

2. Status of Activity: (please check one)

_____ **Ahead of Schedule** _____ **On Schedule** _____**Behind Schedule** __**X**___ **Completed**

Comment:

3. Completed actions, deliverables and results; any major issues or variance between planned and actual activities.

Note: This final small-plot project report is very similar to the March 31, 2014 report previously filed. This is reasonable since the Mixed Model analysis detailed below includes the relevant and available canola small-plot yield data (i.e. 3 years of data, 2011-2013 Canola Performance Trial data, CPT). The CPT canola small-plot data collected during the 2014 growing season was not available in time for analysis and inclusion in this final report. Australian researchers working on Mixed Model analysis of crop variety trial data have indicated that approximately five years of data seems to be optimum for analysis, provided that there is some 'connectivity' in the yearly data – that is, that varieties are not all unique in each year of testing, but rather that there is more than one year of testing for a number of varieties in the multi-year dataset.

As a 'proof of concept' exercise, the former existing small-plot Prairie Canola Variety Test (PCVT) yield data (2003 to 2009 PCVT data, as provided by the Canola Council) was subjected to Mixed Model analysis using the specialized computer program 'ASReml', which is capable of analyzing very large datasets. The PCVT 2003-09 overall combined dataset (combined over years and growing zones – short-season, mid-season, and long-season zones) was approximately 35,000 datalines. However, the 2003 to 2009 PCVT data shared only one variety in common with the 2011-13 CPT small-plot dataset. This is due to the very rapid turnover of canola varieties, with new varieties/hybrids continually being registered and commercialized. Also, there are some differences in field protocol between the CPT and PCVT trials that may have influenced yield (e.g. spraying of herbicides, arrangement of plots). As background information, there were no Canola Council supervised post-registration canola small-plot trials conducted in 2010, hence the gap in the dataset. Therefore, since the PCVT 2003-09 data is not current and may no longer be relevant, the results of the Mixed Model analysis of PCVT data are not presented in this report. If you would like to see these results, please let us know.

Dr. Rale Gjuric of Haplotech, who is co-ordinating the independent canola variety small-plot trials on behalf of the Canola Council, sent the CPT yield dataset (with replicate values) to us each year in November. Each year, the small-plot dataset was appended to the previous years' small-plot dataset to eventually provide three years of data (2011-13) for Mixed Model analysis.

Prior to analysis, each year the CPT small-plot raw dataset was re-formatted for the statistical computer programs, SAS and ASReml. Also, variety and location names were checked for consistent and correct spelling (consistent spelling with the previous years' data). A Mixed Model analysis was conducted on the 2011-13 CPT small-plot yield data using the Mixed Model statistical computer software program ASReml, which has been designed/optimized to accommodate large datasets. The summary tables of the small-plot Mixed Model analysis are attached to our companion CARP report on analysis of canola strip trial data. At the time of initiation of these CARP projects, it may not have been fully appreciated that the 'Strip Trial' project is very comprehensive in that it essentially includes and encompasses this project on small-plot canola data analysis. That is, before a comparison can be made between strip trial and small-plot results, the small-plot data also must be subjected to Mixed Model analysis. The Strip Trial project report includes this comparison to small-plot Mixed Model results, which is why the small-plot results tables are appended to the Strip Trial report (and not duplicated in this report). Grouping the tables together appended to one report facilitates comparison between small-plot and strip trial results, including relative rankings of varieties in each growing environment.

Summary of 2011-13 CPT Small-plot Mixed Model results (refer to tables appended to our CARP Strip Trial report):

1) For the small-plot dataset CPT 2011-13 (inclusive), there is a summary table of canola yield arithmetic means "by Year", "by Province", and" by Year-Zone". The Year 2013 had the highest yield, approximately 1.5 fold higher than Year 2012. For the "by Province" means, B.C. had the highest yield (followed by Alberta), while Manitoba had the lowest yield. This result seems counter-intuitive and probably does not correspond to what might be expected if a long-term (10 year) dataset was available. This is a result of the Short Season Zone (SSZ) having the highest yield in both 2011 and 2012, while the Long Season Zone (LSZ) had the lowest yield in both 2011 and 2012, probably due to growing season weather in the LSZ which was quite hot and dry in both 2011 and 2012 during canola flowering. In 2013, LSZ had the highest yields followed by SSZ. This likely was due to the weather (temperature, precipitation) experienced in general by the geographic zones (refer to a map of canola growing zones - http://www.seed.ab.ca/pdf/seed_winter10_49_63.pdf).

2) For the small-plot results, 'Variance Components' were tabulated. The table of Variance Components details the variability in canola yield associated with various factors/effects in the statistical model. The statistical model used in the analysis can be deduced by the listing of effects and interactions in the Variance Components table. The major effect 'Year', and the interactions of 'Zone by Location' and 'Year by Location' were the important effects and interactions in terms of percentage of total variance; that is, these were the important effects and interactions in explaining the observed variability in canola yield in this dataset. All other effects and interactions (not including error/'Variance'/'Residual') were relatively small in terms of their contribution to total variance. The sum of all effects and interactions which included 'Variety' (genotype) were not very important in terms of percentage of the total variance at 5.4%. This lack of importance of genotype in explaining the variability in yield is similar to other crop variety datasets that we have analyzed using Mixed Model procedures, and is in agreement with the scientific literature. Note that the Zone by Variety interaction

is not an important variance component, which indicates that Variety rankings do not 'flip' significantly by Zone (even though actual average yields in kg/ha differ between Zones). This indicates that presenting a summary of canola variety performance by Zone (as in Seed Manitoba 2014) is not necessary from a statistical point of view (at least for the dataset of 2011-13 small-plot CPT). Note that in terms of statistical model specification the factor "Location" is completely nested within "Zone".
3) For small-plot results for varieties that had two or more years of testing (i.e. a larger number of observations), there was generally good correspondence between arithmetic mean values and BLUP (Best Linear Unbiased Predictor) yield estimates. Based on the matrix mathematics and algorithms underlying Mixed Model analysis, as the number of observations (statistical 'n') becomes large, the arithmetic mean and BLUP estimate will converge to the same value (this is represented by the "U" or "Unbiased" in the BLUP acronym). Where the difference between the arithmetic mean and BLUP estimate was relatively large, these varieties were tested only in one year and as a result had limited data.

4) For small-plot results, based on the BLUP estimates and an approximate LSD value of 175 kg/ha, there are large groupings of canola varieties where yield does not statistically differ. For example, the second highest yielding variety 'L261' (BLUP value) is not statistically different from the 18th highest yielding variety 'CAN 1990'. The approximate LSD value at the 0.05 level of significance is calculated by multiplying the "Overall Standard Error of Difference" by 2 (or 1.96 as per statistical t-table). Refer to our Strip Trial report for a discussion regarding statistical significance of BLUP's (with regard to the trait, yield), and the practical application of this information in selecting crop varieties.

5) To compare CPT 2011-13 small-plot variety yield rankings to CPT 2011-13 strip trial yield rankings it is necessary to express the variety yields as a percent of a designated Check variety (or percent of a median BLUP yield value – the median BLUP value essentially treats the entire dataset as a 'basket' of Checks). This percent of Check approach is necessary for comparison because the overall arithmetic average yield (and yield potential) is quite different between the two growing environments. In this case, the overall arithmetic average small-plot yield was 3553 kg/ha (by Year average), while the strip trial overall average "by Year" was 2773 kg/ha. This is a difference of approximately 30%. Interestingly, in this case the median BLUP yield value is very close to the variety '73-75RR' BLUP yield value – this variety has been used as a check in Seed Manitoba (2013, 2014). Refer to our Strip Trial report for a detailed discussion and comparison of small-plot and strip trial variety yield rankings.

Conclusions: Mixed Model analysis of canola small-plot data is appropriate and provides variety yield estimates that appear to be accurate. Currently, the results of CPT yield data analysis are presented as arithmetic means (in the 'Seed Manitoba' publication). Mathematical and statistical theory indicate that least-squares linear models (which is Mixed Model analysis) will always provide better or equal results to an arithmetic mean based approach. The advantage of Mixed Model analysis and adjusted 'means' (BLUP estimates) over arithmetic means becomes apparent where data are limiting and/or the year (growing season weather which influenced yield) was unusual as compared to a 10-year mean yield. Refer to Dr. Anita Brûlé-Babel's Manitoba Agronomists Conference presentation for a clear example of the superiority of Mixed Model analysis of crop variety trial data (using Spring Wheat as an example) http://umanitoba.ca/faculties/afs/agronomists_conf/media/Brule-Babel_Pres_Dec_13_2012.pdf

4. Significant Progress/Accomplishments
See above (Section 3).
5. Research and Action Plans/Next Steps
The results of the Mixed Model analysis of the 2011-13 CPT small-plot data have already been sent to some key persons involved in the organization of canola small-plot variety trials and presentation of results. We welcome further consultation and discussion on this topic. When the 2014 CPT small-plot yield data becomes available (November, 2014), it can be merged with the 2011-13 CPT dataset (provided that there is some overlap in varieties tested) and an updated Mixed Model analysis conducted. Additional years of data in the Mixed Model analysis should lead to even more accurate variety BLUP yield estimates and variety rankings. The scientific literature suggests that five years of multi-location crop variety data in a Mixed Model analysis is desirable.
6. Budget impacts in the event major issues or variance between planned and actual is noted:
None anticipated.

Please forward an electronic copy of this completed document to:

Gail M. Hoskins
CARP Coordinator
Canola Council of Canada
400 – 167 Lombard Ave.
Winnipeg, MB R3B 0T6
Phone: (204) 982-2102
Fax: (204) 942-1841
E-Mail: hoskinsg@canolacouncil.org